Visual Affordances in Natural User Interfaces

By

Omprakash Rudhru

Submitted in partial fulfillment of the requirements for the degree of

Master of Human Interface Technology

At

The Human Interface Technology Lab New Zealand

College of Engineering

University of Canterbury

12 July 2016

Supervisor: **Dr. Gun Lee**
Co-Supervisor: **Prof. Mark Billinghurst**

Acknowledgements

I express my deepest gratitude to my supervisor, Dr. Gun Lee for the innumerable number of hours spent on guiding me throughout my thesis. I would also like to thank Prof. Mark Billinghurst for his guidance and Prof. Rob Lindeman for his valuable inputs on my prototype design. To my course professors for preparing us for the long road of thesis, Assoc. Professor Christoph Bartneck and Dr. Adrian Clarke, I give my thanks. Finally, I am deeply thankful to all my family, friends and classmates at the HIT Lab NZ for their patience, guidance and support during each stage of my Master's thesis.

Abstract

As part of my Master’s thesis, I conducted research on the usage of UI agents in gestural interfaces. The research focused on providing visual affordances through UI agents for large-screen displays. User engagement is critical for many public information systems or large-screen displays. The gestures used for the interface need to be easy to understand. Previous research has shown that users need feedback for understanding natural gesture interaction. To achieve the goals of the thesis, I built three prototypes in an iterative model. These explore different ways of using UI agents in providing visual cues. A focus group and two user studies were conducted to test the prototypes. Prototypes were evaluated based on initial user engagement and system usability on the main interaction phases. Results of the user studies show that using UI agents as visual affordances is more engaging and results in fewer errors during gesture interaction. The success of the UI agent depends on its relationship with the interface design. Overall, UI agents are effective in giving users feedback in order to help them understand the interface. These findings are important for designing public information systems where user engagement is required.

Contents

Figures

Tables

1. Introduction

Augmented Reality [1] (AR) has expanded over the past three decades from a field exclusive to scientific laboratories, to personal handheld mobile devices. Various sectors have adopted AR to enrich the real world by superimposing animated 3D virtual objects over real objects. This sparked research in the different interaction methods that can be used to access the information. For example, Augmented Reality can be used with large-screen display systems in shopping, which allows people to use gestures to virtually try on clothes.

Though gestures and voice-based interaction are considered the most natural ways to interact, they have a steep learning curve. Visual affordances, a design feature that provides visual cues to direct users to take certain action, is useful in overcoming this learning curve. Among different types of visual affordances, cognitive affordances in particular provide visual cues about the interface on how to interact with the system. Much research has been carried out to explore visual affordances in gesture interfaces. For instance, to improve the learning curve, combinations of ghosting [2], video[3], light controls [4] and traces of gestures [5] have been explored. Previous research in these areas showed that visual affordances can help users learn the interface, but has found no significant improvements in the initial learning curve.

With regard to visual affordances, little to no research has been done to using user interface (UI) agents. The use of earlier intelligent forms of UI agents like Microsoft Clippy[1] has decreased over time. With regards to gesture interfaces, more research has been done in the area of using either UI agents for human-to-human interaction [6] or UI agent-to-human interaction [7] to accomplish tasks. Another such example is using UI agents for voice-based interaction like Apple Siri[2] or Microsoft Cortana[3]. Earlier types of visual affordances for gestural interactions used other cues, but these visual cues were not successful in helping the user intuitively learn the interface system. There is little or no active research in the area of using UI agents for visual affordances.

Visual affordances for various gestures using UI agents are expected to allow the user to intuitively learn the interface system. This research has direct application to public information displays such as virtual shopping, public kiosks and advertising campaigns. In these displays,

[1] https://www.theguardian.com/media/2001/apr/11/advertising2
[2] http://www.apple.com/nz/ios/siri/
[3] https://www.microsoft.com/en/mobile/experiences/cortana/

initial engagement and intuitiveness of the interface system are key factors in the successful adoption of such interfaces. In addition, this system is expected to help first-time users of gestural interfaces.

For the master’s thesis, I have explored a combination of visual affordances and UI agents for gesture based interaction with public information systems. Affordances have been explored in the context of using visual cues to interact with gestural interfaces. It has been demonstrated that visual cues can reduce the learning curve.

A combination of UI agents and visual affordances will provide numerous possibilities that could improve the gesture based interaction with the public information systems. Further research can be carried out to find the impact of different animation characters used as UI agents and their relation to the interface.

1.1. Research Questions

Within the context of gestural interactions with public information displays, we attempted to answer the following research questions:

- How will visual affordances using UI agents reduce the learning curve for interactions? My hypothesis is that the proposed natural gestural system will help to reduce the learning curve.
- Can the usage of visual affordances using UI agents help improving the user engagement during interactions?
  The goal of this research question is to evaluate if users are able to complete their tasks using the proposed interface in a more engaging way.
- How will visual affordances using UI agents help interaction with public information displays?
- Is there any impact of types of UI agents on the performance of user during interactions? This research question is aimed to evaluate if using different types of characters as UI agents have any impact on the user performance while performing different tasks.
- Can the usage of animation characters as UI agents help to encourage the users to take initiative in interacting with public information displays?

I will be testing the hypothesis, which is that the type of animation characters as a UI agent would have an impact on users to take initiative to interact with public information displays. For this, UI agents are to be designed in such a way that they are always on the screen showing visual affordances.

1.2. Research Scope

The aim of the research is to explore the usage of UI agents to provide visuals cues in gestural interaction with large-screen displays. Research focuses on how UI agents can be used to increase the user engagement while exploring the usability, learnability and accuracy while doing gestures as measurements.

This research covered the type of UI agents that can be used, different engagement techniques like providing occlusion for the UI agent or for the visual cues, and using UI agents as a cursor or buttons. Also, the affordances to help with "Hover to select" gesture is explored as part of this[4].

However, this research did not cover using UI agents in interfaces with more than two choices for user to select. In addition, usage of other gestures besides "Hover to select" are not covered.

Findings from this research can be applied to make the gestural interfaces more engaging for the users. This is especially important in public information systems where initial engagement is an important factor in the success of the gestural interface.

1.3. Thesis Structure

This section describes the structure of the thesis report, as follows:

- Chapter 2: Background research conducted so far in the areas of AR, AR for large-screen displays, visual affordances, and UI agents.
- Chapter 3: The concept design, different service scenarios, and the design thinking behind the prototypes that were built during this research.
- Chapter 4: Proposed design ideas.

[4] https://support.xbox.com/en-NZ/xbox-360/accessories/body-controller

- Chapter 5: Prototype version one, design, hardware and software specifications, and the implementation details.
- Chapter 6: Focus group study and discussion of the results.
- Chapter 7: Prototype version two with the design and implementation details.
- Chapter 8: Evaluation of prototype version two and discussion of the results.
- Chapter 9: Prototype version three, with the design and implementation details.
- Chapter 10: Evaluation details of prototype version three and discussion of the results.
- Chapter 11: Limitations of this research study.
- Chapter 12: Future research opportunities.

2. Background Research

In order to review state-of-the-art technology, related work was investigated in the academic and industrial field in using visual affordance cues for natural gesture interaction with large-screen displays. This chapter summarizes the related work by category. The first section summarizes the existing literature on augmented reality, the concept of affordance and how they are used in user interface design. The next section analyses prior work on using visual cues to help users learn and use gesture interfaces. This chapter also reviews software UI agents as one of the potential visual cues to help users interact with large-screen displays using motion gestures.

2.1. Augmented Reality

Azuma et al. proposed that Augmented Reality is super-imposing virtual objects upon the real world [8]. It is a variation of Virtual Reality wherein the key difference is that AR supplements reality, whereas Virtual Reality completely replaces it. Though initially it was only about augmentation of vision and sound, it later extended the definition to apply to all senses by using sensory substitution. For example, visual cues can augment hearing and auditory cues can augment vision. Touch and smell sensations can also be substituted using sensory cues [9]. Augmented Reality appeared for the first time in 1950s through Morton Heilig's "Sensorama" [10], a prototype of a multisensory cinematic experience (see Figure 2.1). This prototype draws the user into the movie by making use of multiple senses. The next breakthrough in AR was the invention of a see-through head-mounted display (HMD) by Ivan Sutherland in 1968 [11]. It was limited in graphical quality, heavy in weight and had to be suspended from the ceiling (see Figure 2.2).

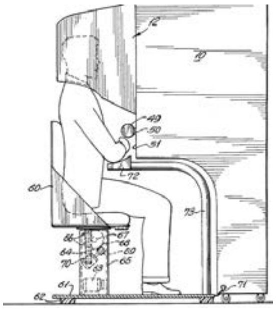

Figure 2.1: The Sensorama, from U.S. Patent #3050870 [10]

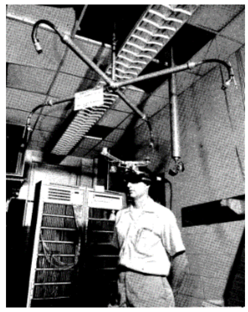

Figure 2.2: Head mounted display [11]

Though Augmented Reality was around in some form, the phrase was first proposed by Professor Caudell in 1990 while working on HMDs to be used to aid the human side of aircraft manufacturing operations [12]. Caudell's prototype (see Figure 2.3) successfully overlaid diagrams onto a real-world position based on the user's head position. This was the first demonstration of AR using real-world positions.

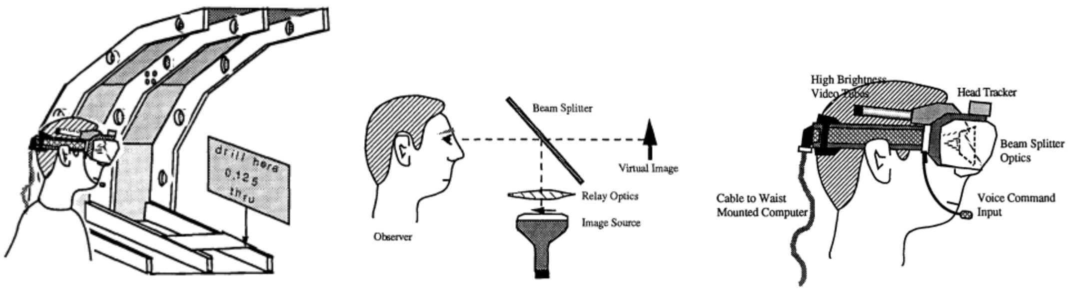

Figure 2.3: Explaining different principles of head-mounted display with Augmented Reality [12]

Around the same time two major developments were made in Augmented Reality. First, Rosenburg et al. [13] developed the first fully-functional AR. Known as virtual fixtures, this development in AR demonstrated that AR can enhance people's performance on tasks. Second, Feiner et al. [14] presented a major paper on KARMA, an AR system which explained how to maintain a laser printer using a HMD. In 1997, Azuma et al. [8] wrote the first comprehensive survey of AR, a definition that was widely acknowledged.

Until 1999, AR was something that was not yet accessible to common users. The hardware was bulky and the software, complex. This changed with the open-source version of ARToolKit, developed by Kato et al. [15] at the Human Interface Technology (HIT) Lab. The ARToolkit combined live video with virtual object interaction using 3D graphics. Able to run on any platform, the toolkit was a major milestone in bringing AR to developers.

Soon after ARToolKit, first AR video game was showcased by Thomas et al. [16] in 2000. Used outdoors, it involved GPS sensors, gyroscopes, head-mounted displays and computer backpacks. From 2008, as the hardware of mobile phones improved, and as camera sensors and processing speeds became more powerful, AR tracking methods also improved.

Today, the Microsoft Kinect is one of the most promising natural user-interaction methods (reference). As it lets users interact by using their natural body positions, this makes it promising for large-screen displays. With the Kinect, users no longer interact using a pointing device, such as a mouse. Thus, visual feedback is missing in this interface [17].

2.2. Augmented Reality in magic mirror-like systems

Many applications have begun to use Augmented Reality in magic mirror-like systems. These systems mimic a mirror with extra features by using Kinect to overlay virtual 3D images over the real-life image, usually using a large-screen that is inverted to resemble a mirror. People who use this "magic mirror" would then see, on the screen, an augmented moving live image of themselves, tricking them into thinking that what they see on the screen is real.

Figure 2.4: Mirracle interface [18]

One of the systems that feature a magic mirror system is Mirracle. Developed by Blum et al. [18], Mirracle features a magic mirror-like system for human anatomy using augmented reality. Using a circular window over the user's body, it overlays 3D images of CT volume scans by tracking the user's pose (see Figure 2.4). Mirracle thereby gives the user information about anatomy. The CT information imagery is done through 3D gestures using Kinect. One of the problems with the 3D gestures is that they do not provide any feedback for zooming or selecting an item, as in multi-touch systems (e.g., cell phones). To overcome this, the developers added a virtual pane of frosted glass. . When the user wants to zoom, they use both virtual hands to approach the "glass" (see Figure 2.5). This addition uses a depth camera to modify the visibility of hands and body based on the user's distance to the virtual interaction plane. This mode of interaction is used to display text for interaction options. They found that the frosted-glass interaction helped the user understand spatial relations between the virtual interaction plane and his body. Children in particular received this addition positively.

Figure 2.5: Frosting Glass Interaction [18]

Another system that makes use of the magic-mirror system is Cisco StyleMe developed by Cisco IBSG [19]. Upon the successful announcement of the concept of "mashopping", Cisco StyleMe was developed to give customers a fun and interactive way to try clothing and accessories virtually without having to physically try them on. It consists of a life-size mirror-like screen that overlays the users' image with images of clothes. Shoppers can create outfits quickly and easily by using gestures and touch (see Figure 2.6).

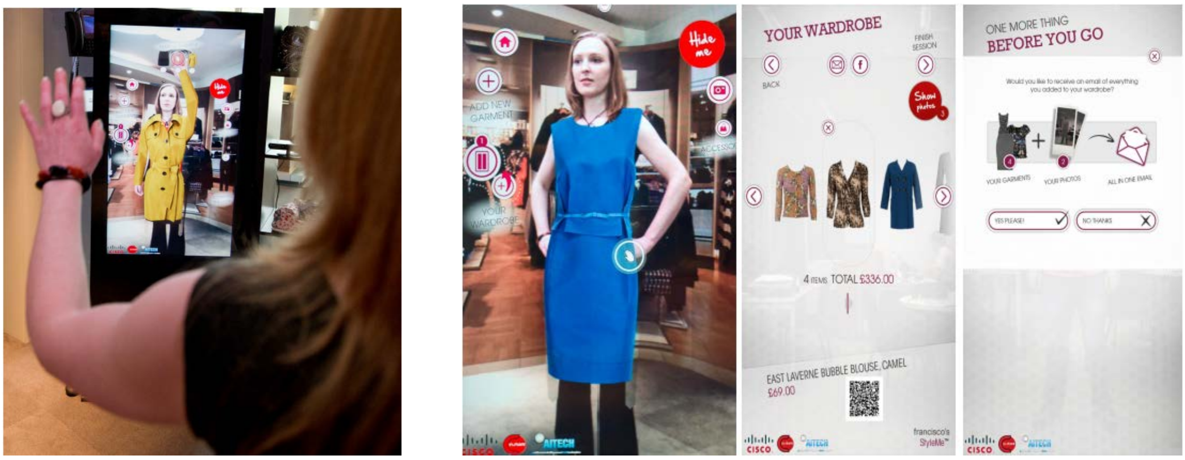

Figure 2.6 Cisco StyleMe Interface Source: Cisco IBSG, 2011 [19]

2.3. Affordance – Concepts and Background

The concept of 'affordance' in design was first introduced by Donald Norman in his book titled *Design of Everyday Things* [20], where it was described as a design feature that affords (i.e., induces or encourages) the user to make a certain action while using it. For instance, Figure 2.7 shows various design of door handles on a car that affords different ways of grabbing and opening the door. The book introduced various design principles (including conceptual model, constraint, and affordance) and their psychological explanation. Based on industrial design, these principles are summarized by the concept of human-centered design.

Figure 2.7: Car door handles [21]

Since its introduction to the design community, the concept of affordance has evolved due to further investigation and especially application to the field of Human-Computer Interaction (HCI) design. Captelinin [21] summarizes the history and concept of 'affordance' under the context of Human-Computer Interaction design, providing many references to related work on the concept of affordance and research in HCI.

It is easy to fall into the trap of misusing the term (Norman, 1999, as cited in *Interaction Design*), as frequently people do not understand the difference between perceived and real affordances. Virtual user interfaces do not have real affordances, but are better conceptualized as perceived affordances, which are in essence learned conventions.

This research project focuses on the concept of cognitive affordance [22], which is more relevant to graphical user interface design. With cognitive affordance cues users perceive certain hints of how to interact with the interface through recognizing the visual cues provided on the screen.

2.4. Visual Cues and Guides for Gesture Interaction

Together with speech recognition based natural language interfaces, gesture interaction is widely recognized as one of the most natural way of interacting with computer. However, Quek et al. [17] pointed out that both speech and gesture based interfaces are not as intuitive as they are thought of as they have a very steep learning curve where users have to learn the vocabulary of words or gestures that the system can recognize.

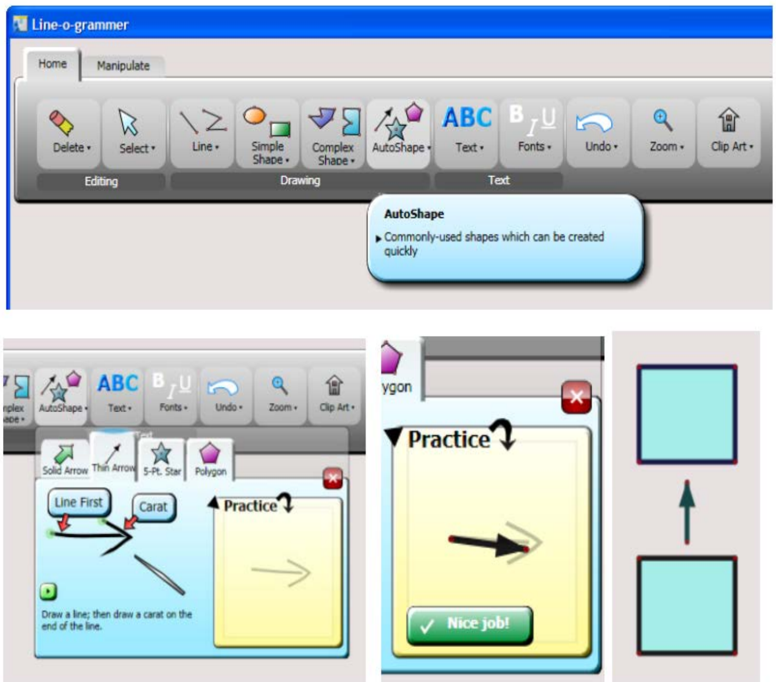

Figure 2.8: GestureBar interface [23]

To overcome this limitation, there have been many research investigations on providing visual cues and guides for helping users learn and use gesture based interfaces. For instance with 2D GUI based mouse gesture interface, Bragdon et al. [23] proposed to leverage a common GUI toolbar, but in place of executing commands, richly discloses how to execute commands with gestures, through animated images, detail tips and an out-of-document practice area (See Figure 2.8). As a more dynamic guidance tool for gesture interaction, Bau and Mackay [5] proposed an interface named OctoPocus, an animated visual guide that shows optional gesture paths to trace to help learn gestures in 2D GUI. With this interface, the optional gesture paths appear and disappear as the user traces them. (See Figure 2.9). Compared to conventional Help menus, OctoPocus helped users better remember and execute gesture-based command sets, as well as showing a marked improvement in learning. Users also showed a preference for learning of gestures with Octopus as compared to the original interface.

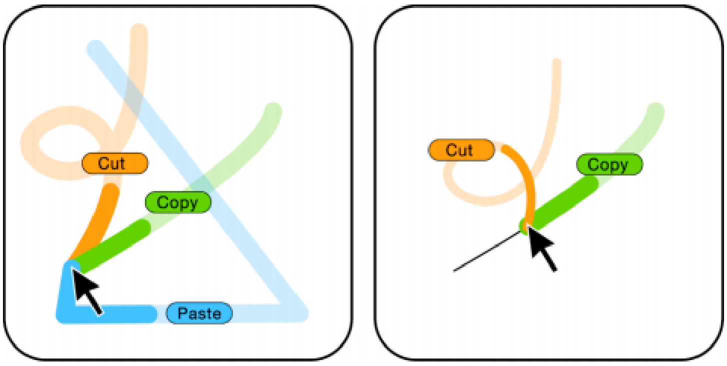

Figure 2.9: OctoPocus interface [5]

For multi-touch and whole-hand gesture interfaces, Freeman et al. [3] proposed ShadowGuides which provides on-demand assistance to the user by combining visualizations of the user's current hand posture as interpreted by the system (feedback) and available postures and completion paths necessary to finish the gesture (feedforward) (see Figure 2.10). Users learning with ShadowGuides were found to remember more gestures and had higher preference for the help system as compared to users learning with video-based instruction. This can be attributed to the continuous feedback and feedforward features which aided in error correction. However, the authors acknowledged that although ShadowGuides allows for attention to be focused on specific important features of the gesture, it pales in comparison to video-based instruction in conveying the general gist of the motion.

Figure 2.10: ShadowGuides interface [3]

As the number of research work on gesture guidance grew with various approaches, Delamare et al. [24] proposed and created a web based tool that provides summary, references, and tools to explore design space for gesture guiding systems not only for 2D GUI interfaces, but also including 3D (three-dimensional) motion tracking interfaces. The tool allows easy access to useful description of systems, comparison of two or more guiding systems, finding areas of the design space that were previously unexplored, as well as provides a webpage for designing guiding systems. These greatly enhance knowledge capitalization of state-of-the-art guiding systems for researchers and practitioners.

Sodhi et al. [4] proposed LightGuide, a gesture guidance system for 3D motion tracking systems that projects guidance hints directly on a user's body (see Figure 2.11). These projected hints guide the user in completing the desired motion with their body part. In a user study conducted, LightGuide users were found to perform with a mean error of 21.6mm, 85% more accurately than when guided by video. However, the authors noted that there is a trade-off in speed. Similar to Freeman et al. [3], they attributed the slowdown in speed to the inability of users to see the general gist of the motion before performing the gesture. It was therefore suggested that gesture guidance systems be combined with video to attain maximum efficiency in learning accuracy and speed.

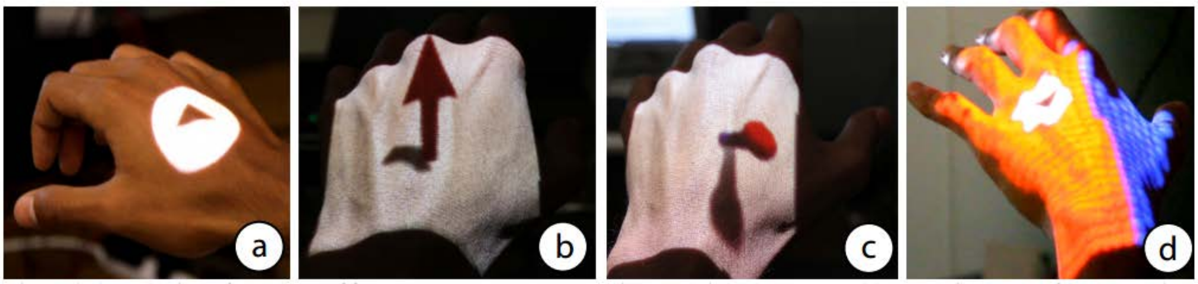

Figure 2.11: LightGuide interface [4]

For Augmented Reality gesture interfaces, White et al. [2] proposed various types of visual hints for Tangible AR systems, guiding potential actions and their consequences, enabling discovery, learning, and completion of gestures and manipulation (see Figure 2.12). It was found that combinations of ghosting with text or animation were the most preferred.

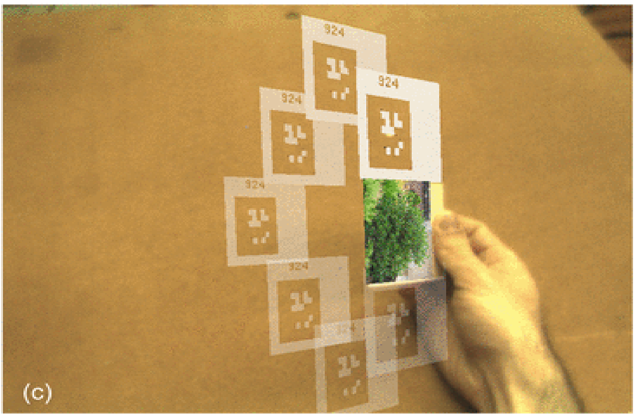

Figure 2.12: Visual hints for gesture interaction in Tangible AR [2]

Anderson et al. [25] proposed an AR mirror based motion training system named YouMove that allows users to record and learn body motions for training purpose. The system trains the user through a series of stages (posture guides and movement guides) that gradually reduce the user's reliance on guidance and feedback (see Figure 2.13). Results from the user study found that YouMove improved short-term retention and learning by more than a factor of 2 as compared to video demonstrations. In parallel with previous systems (ShadowGuides and LightGuide), it was found that training with the YouMove system took longer than with video.

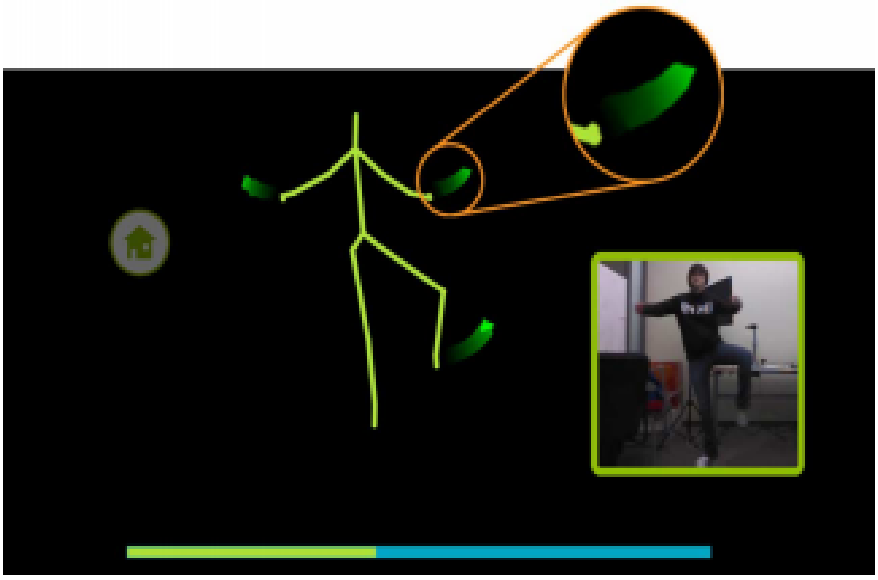

Figure 2.13: YouMove interface [25]

2.5. Software UI Agents and Gesture Interaction

Software user interface agents are virtual characters that work as a part of human-computer interface. Laurel [26] explains the concept and usefulness of interface agents based on the concept of anthropomorphism, while Lieberman and Selker [27] provides more comprehensive introduction to the concept and researches on UI agents. With its unique feature of autonomy and user friendly appearance, UI agents have been considered as one of the representative human-computer interaction paradigm together with more tradition direct manipulation in graphical user interfaces [28].

Not only as a mere fancy representation of a virtual character, UI agents are considered as intelligent autonomous virtual assistants that human users can delegate their work on, collaborate with, or even play with [29] [30] [31]. Researchers tried applying UI agents to various applications and tasks including web browsing [32], presentation [33] (see Figure 2.14), and education [34] (see Figure 2.15). With growing interest and application of UI agents, Microsoft built a software framework for supporting interface agents on Windows platform. Efforts of Microsoft on using agents as user interface have been reflected in their products, such as ‘At Home with Bob’ and ‘Clippy’ UI agent in Microsoft Office products (see Figure 2.16).

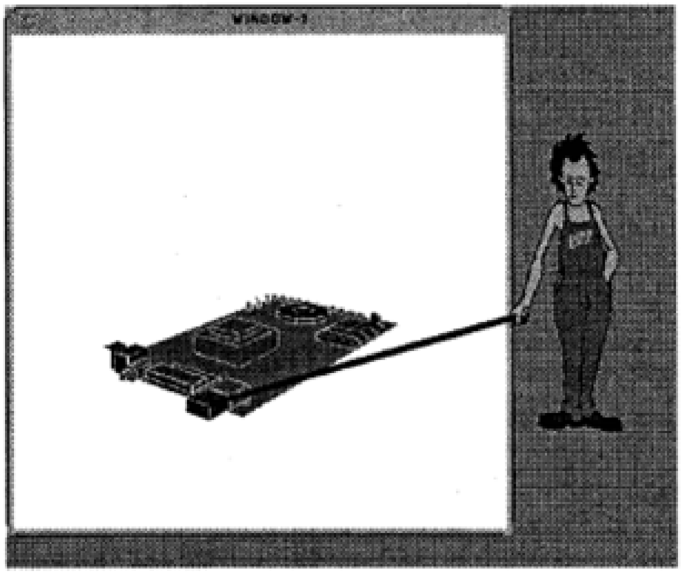

Figure 2.14: UI agent for presentation [33]

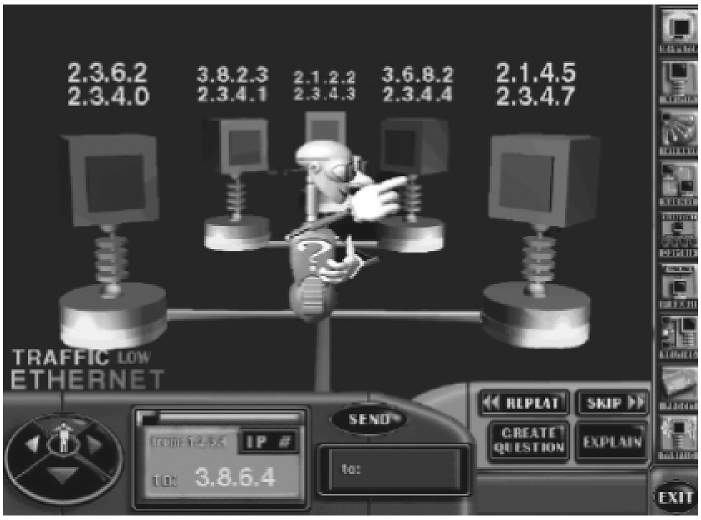

Figure 2.15: Cosmo, a pedagogical [34]

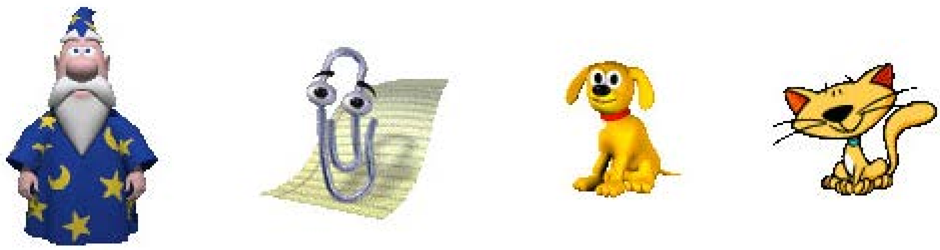

Figure 2.16: Various virtual characters supported in Microsoft interface agent framework

While the interest in UI agents with animated virtual characters have been falling, the concept of UI agent has evolved into a form where emphasizing on the aspect of UI agents as intelligent virtual assistants that users can talk to using natural language. Apple's Siri on iOS mobile operating system[5] and Microsoft's Cortana are typical examples of UI agents that users can interact through speech recognition or simple text input.

While natural language has been considered as the ultimate method for interacting with UI agents, many researchers also investigated how communication between the user and the UI agent can be improved by adding other modalities. Gestures have been actively investigated as a complementary communication channel to the natural language based interaction between user and UI agent, just as in human-human communication. Many researchers have been focusing on animating facial and body gestures of UI agents to convey subtle nuances such as emotion when a UI agent talking to the user [35] [36] [37] [6] [38], while some researchers also looked into letting UI agents to recognize the user's body gestures to interact with. For

[5] http://www.apple.com/nz/ios/

instance, Maes et al [7] investigated using computer vision technique to recognize users' body motion to let them interact with a dog-like virtual agent (see Figure 2.17) while Cassell et al. [39] investigated on supporting more human-to-human communication like interaction with an interface agent (see Figure 2.18).

Figure 2.17: The ALIVE system [7]

Figure 2.18: Rea, a UI agent [6]

While many researchers looked into gesture based interaction as a communication channel for interaction between UI agents and users and UI agents helping users for various tasks, to our best knowledge, there have been no prior work investigating UI agents for helping and guiding users to interact with the system using gesture interaction. This research investigates how UI agents can be used for providing visual cues and guidance to the users interact with a public information system using gesture interface.

3. Concept Design

The main purpose of the system we are investigating in this project is to provide interactive information displays in public spaces. In this chapter we first describe the typical system setup and service scenarios in a couple of representative business environment. Next we describe the use case scenarios in details that illustrate the proposed idea of using software UI agents for helping and guiding public users to interact with the system using body gestures.

3.1. Service Scenario

The overall service scenario is illustrated in Figure 3.1 with a typical system setup. The basic system configuration includes a large-screen display installed in public environment (e.g. shopping malls, libraries, airport, etc.), being accessible to public users, information through a server computer connected to the Internet. The main method to interact with the display is through motion gesture. To make the display interactive to the public users, the system includes a set of motion sensing devices, such as depth sensors or a video camera.

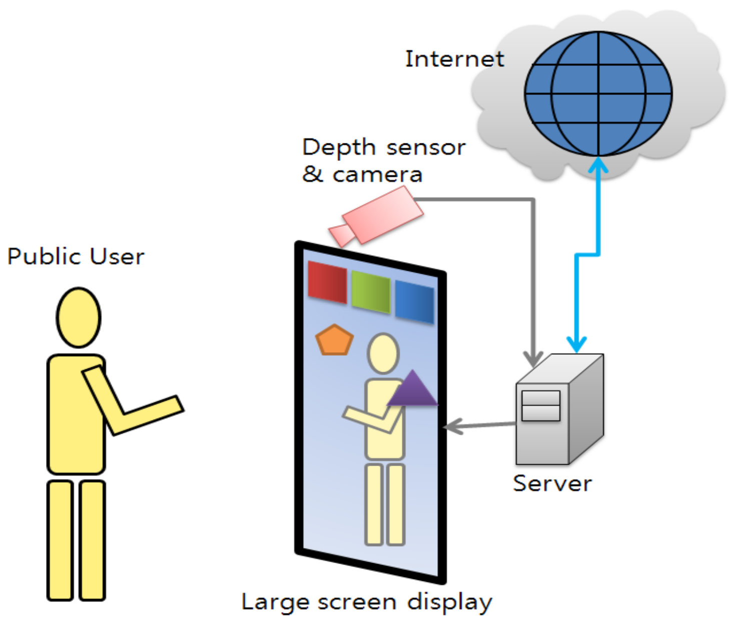

Figure 3.1: Typical system setup for the service scenario

While the system can use traditional 2D and 3D graphics for visualization, in this research project we focus on the metaphor of virtual mirror. The camera used for motion recognition can provide the video stream as a background for the visualization, while the computer graphics including GUI (graphical user interface) elements are overlaid on top of this video background. This creates an experience for the users as if they are standing in front of a mirror overlaid with computer generated graphical content.

The system setup described above can be used in various service scenarios, including public information services, entertainment, and commercial applications. For instance, the system can provide guiding information in public places, such as shopping malls, libraries, or airports. The large-screen display can display a variety of information including maps, directions, business hours, or special events. The system can also be used for a more commercial application where it provides entertaining interactive content to attract public users that leads into shopping or advertisement. For example, the system could overlay virtual objects on the real world, virtually reflected on the display, creating an augmented reality experience. This feature can be leveraged to visualize products as in real use. For instance, users can virtually try on apparels or other wearable products while standing in front of the virtual mirror. Other types of product can appear on the floor in front of the user or on the wall behind. Such an experience could be extended into online shopping by integrating with a mobile commerce system.

3.2. Use Case Scenario

Based on the service scenario described in the previous section, here we add more details on how users would interact with the system using their body motion and gestures. As discussed in the literature review, we especially focus on how software UI agents can be applied for engaging with and guiding the users given that most of the users will be the first time users, or having not much experience with the system.

The very first step in the use case scenario is the system waiting for a user to engage with it. Many public interactive kiosks spend this idle stage by showing title screen or advertisement. In addition to showing title screen or advertisement and passively waiting for a user to engage, the proposed system can actively attract users by tracking potential users in front of it and interactively respond to the motions of the people passing by.

Figure 3.2 shows an example of how the proposed system could actively attract people by using a software UI agent responding to potential user's motion. While the system is showing a title screen or an advertisement, as people pass by in front of the system the system tracks the closest person as a potential user to create a cut out portion on the title screen and show his or her head. In addition, a software UI agent, such as a virtual butterfly, can be animated to follow the potential user, trying to grab his/her attention. This behavior of the system could not only gain attention of potential users but also reveal that the system can track user's motion and suggest using gestures to interact with it.

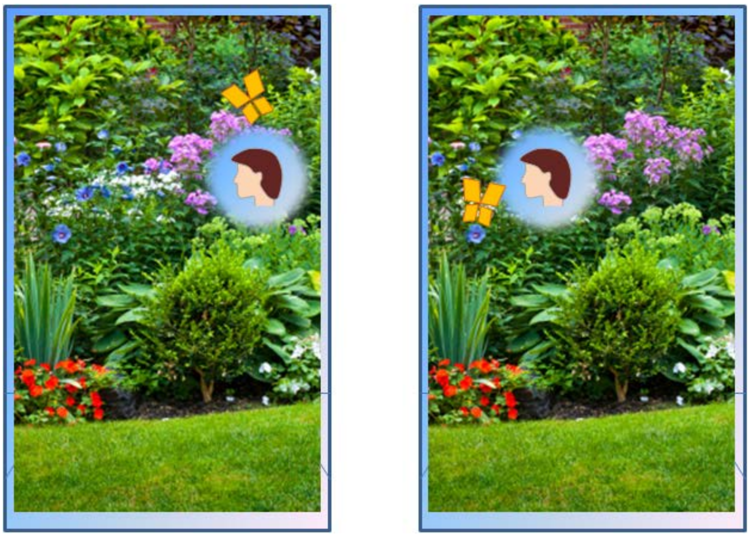

Figure 3.2: Use case scenario – Butterfly Scene #1: Idle

Once the potential user shows his or her interest by slowing down or facing towards the screen, the system can further guide the user to be ready for using the system. Figure 3.3 shows an example of such an engagement stage where the UI agent guides the user to move to an ideal location where user can stand for proper interaction with the system. For instance the butterfly which was following the user can start circling around between the user's current and desired position to stand at. In this way, users can understand the system is recognizing their interest on the system and also learn the basic behaviors of the UI agent and how to interact with it. Once the user comes to the desired place to fully interact with the system, the screen stops showing the title screen (or advertisement) and proceeds into the main interaction screen (e.g. virtual mirror).

Figure 3.3: Use case scenario – Butterfly Scene #2: Engagement

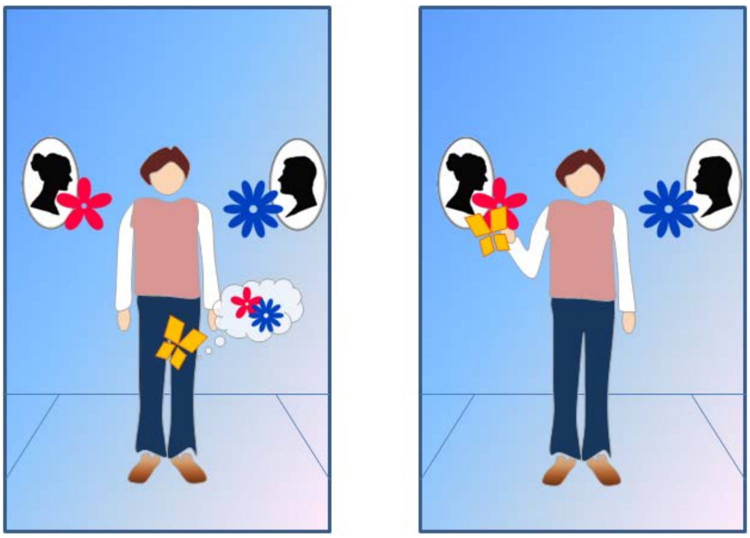

Figure 3.4: Use case scenario – Butterfly Scene #3: Gesture guidance

Figure 3.4 shows an example of how the UI agent can guide a user to select options on the screen using gestures. For instance, when a user has to choose between two options (e.g. gender), the buttons for each option can be shown on each side of the user. As the UI agent follows the user's hand, the user can perceive that he or she should be using their hand to point at the options. If the user does not select one of the options for a while, the UI agent can give more hints by either roaming around the user's hand and the two choices as if it is looking for one of the choices. More explicit way to guide the user could be showing a thinking bubble of the UI agent implying that it is looking for one of the choices. To make this scenario more compelling, the menu items (or buttons) could be designed to have a relationship with the UI agent character. For instance, for a butterfly flowers can be used to represent the menu items.

To prevent accidental operation, many gesture based interfaces require users to hold their hands pointing at a UI item (e.g. a button) to confirm their choice. Common way to indicate that the user has to hold on to confirm is showing a timer when the user points at a button. While this provides an explicit cue, it can be easily missed at the first attempt as the user quickly moves away from the button. UI agents can help improve this problem by providing subtle and implicit hints in advance. For instance, as the user's hand approaches one of the options the UI agent follows his/her hand with certain amounts of delay, implying that the user needs to hold his/her hand in place to confirm the choice until the UI agent arrives at the item.

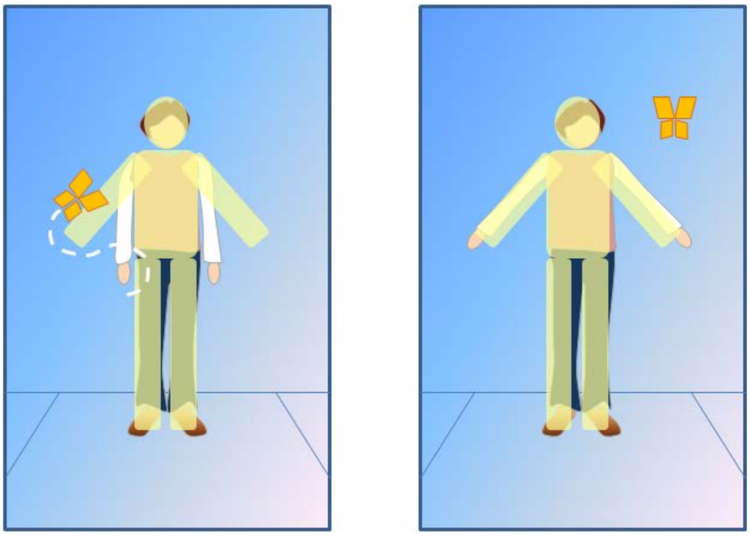

Figure 3.5: Use case scenario – Butterfly Scene #4: Posture guidance

Besides helping users to learn gestures and give hints on how to interact with the system, UI agents can be also useful for guiding users to pose in certain postures as needed. Figure 3.5 shows an example of a UI agent in a butterfly character guiding the user to raise their arms to pose as the yellow outline by roving around between the user hand's current position and the desired position. This can be further extended by introducing multiple UI agents guiding each end of the user's limb for more complex postures, or even combining with more explicit visual cues such as arrows.

The level of expression a UI agent can make depends on the shape and structure of the virtual character used. For instance with the poster guidance, a more anthropomorphic character could actually pose in certain ways to imply the user to follow its motion. To further develop our idea of using UI agents for guiding gesture interaction, we created another scenario using a different virtual character as a UI agent. Here we chose a goose as a virtual character for the UI agent.

While even more human-like characters are available (e.g. a monkey or even human characters), here a goose was chosen for this study to show that full anthropomorphism is not necessary to use in our scenario. Also, as discussed in the literature review, we think the more the character becomes anthropomorphic, the expectations of the user on the character's behavior and communication become higher which can lead into disappointment and disengaging experience.

Figure 3.6 shows the state when the system is idle and waiting for a user to get engaged with the system. The system could be showing an advertisement but leaving the lower portion of the screen uncovered so that it can reflect peoples legs as a virtual mirror. The UI agent in a goose character can appear on this scene following and trying to peck on user's legs to attract them to be engaged with the system. In addition to showing an animation of the UI agent, the system can also use auditory cues to grab more attention such as the goose making noise to grab attention from people passing by.

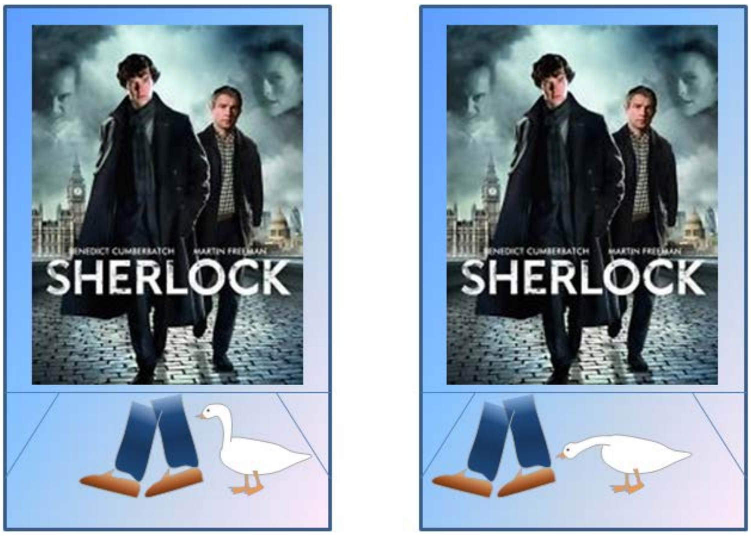

Figure 3.6: Use case scenario – Goose Scene #1: Idle

Once the potential user notices the goose and faces towards the screen, the goose can guide the person to stand at the ideal position to interact with the system. Figure 3.7 shows the goose pointing at the place marked with a red circle suggesting the potential user to stand there. Once the user stands at the right spot, the advertisement disappears and the screen turns into a full virtual mirror scene for further interaction.

Figure 3.7: Use case scenario – Goose Scene #2: Engagement

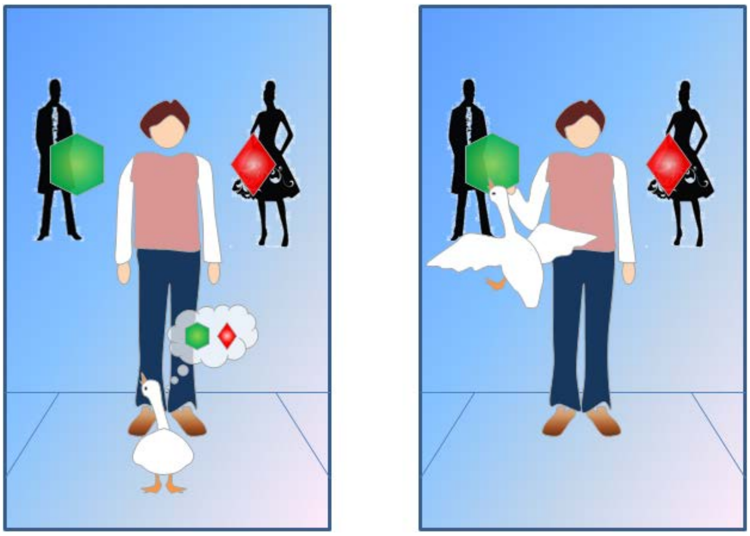

Figure 3.8: Use case scenario – Goose Scene #3: Gesture guidance

Figure 3.8 shows the scene where the goose UI agent can guide gesture interaction for choosing options. Similar to the case using a butterfly as a UI agent virtual character, the goose can turn its head to follow the user's hand, suggesting it is looking at where his/her hand is pointing at. If the user stands still, the goose can be more active for encouraging the user to move by hopping and trying to reach on the user's hand, looking at the UI items to choose, or even showing a thinking bubble as it was with the scenario with a butterfly. Preventing accidental selection could be also implemented similar to the butterfly's case such as the goose trying to jump and reach the buttons pointed at by the user, suggesting the user to hold their hands on the item until the goose reaches it.

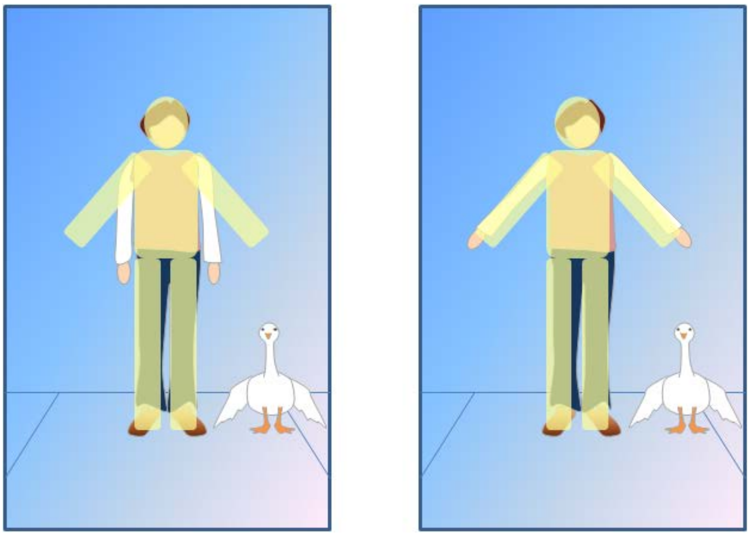

Figure 3.9: Use case scenario – Goose Scene #4: Posture guidance

Figure 3.9 shows an example of the goose UI agent guiding the user to raise his/her hands. As the goose has limbs (wings), it can easily make poses to show the user how to pose, and also add more motions to correct user’s poses. For instance, if the user’s arms are too low, the goose could further raise its arms to emphasize that the user needs to raise their arms more. Or if needed, even make more active motion such as jumping and trying to push the user’s hand higher.

4. Prototype System v1

As a proof of concept and for use in a focus group user study, we designed and implemented a prototype system to show the potential usefulness of the proposed interactive scenarios of using UI agents as guides for gesture interaction. As the first step, we decided to first design and implement a prototype that shows how UI agents can be used to guide users to move their hands to point on UI items (e.g. buttons) and to hover their hand on a button to select.

4.1. Design

We designed a behavior model of the UI agent to follow user's hand as shown in Figure 4.1 where $\mathbf{A}_t$ represents the current position for the UI agent, and $\mathbf{T}$ represents the tracked hand position. Based on the two positions $\mathbf{A}_t$ and $\mathbf{T}$, we can calculate the position of the UI agent in the next frame $\mathbf{A}_{t+1}$ using the following equation:

$$\mathbf{A}_{t+1} = \mathbf{A}_t + s\mathbf{D}$$

where $s\mathbf{D}$ is a scaled vector of $\mathbf{D}$ which is different between $\mathbf{T}$ and $\mathbf{A}_t$ with a scale factor s. The scale factor s is calculated based on the following equation:

$$s = (1/fps) / t$$

where t is the preset amount of time delay in seconds for the UI agent to reach $\mathbf{T}$, and fps is the framerate (frames per second) of the system simulation/rendering loop.

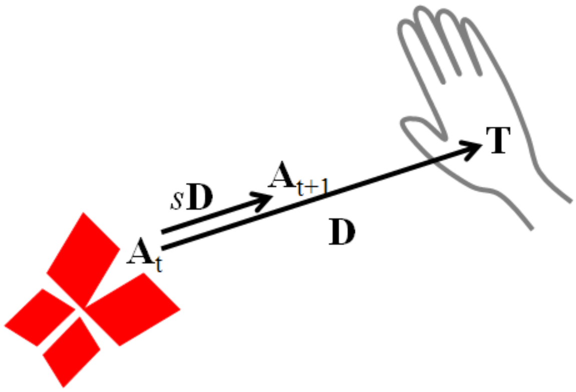

Figure 4.1 Behavior of a UI agent following the user's hand

For guiding the UI agent to face towards a certain direction, we use a normalized vector of $\mathbf{D}$ to set the orientation of the UI agent. The time delay t is normally set to 1 second when the user hand is not on the button. When the user's hand is on the button, *t* is set to the amount of time left until the button gets triggered. In this way we can control the speed of the UI agent to let it arrive at the target position (on the button) just when the button gets triggered.

4.2. Development Environment Setup

We set up the hardware and software environment for the project, and used for development and testing of the UI concepts proposed in the project.

4.2.1. Test Bed Hardware

Figure 4.2 shows the test bed hardware set up for the project. The hardware system setup is formed to provide virtual mirror style visualization. The system has a 52 inch full HD TV (television) screen set up in a portrait pose. The screen shows the graphics generated by a desktop PC (personal computer) running software developed for the project.

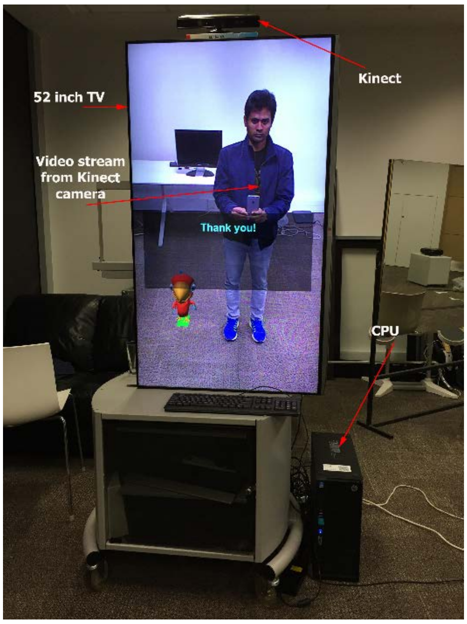

Figure 4.2: Test bed hardware for development

The test bed is equipped with two imaging sensors: a web cam and a Microsoft Kinect depth sensing camera. On the screen, the software presents video image of the user's environment acquired by these imaging sensor, mimicking physical mirror in a virtual way. The video image on the screen is mirrored so that the image is perceived in a similar way the physical mirror works. The Microsoft Kinect sensor is the main imaging device that is used for tracking users' motion and gesture recognition. While the Kinect sensor also has a RGB color camera, the resolution of the video stream has low quality to fit in a portrait oriented screen. To use higher resolution image as a video background on the screen, the web cam can be optionally used with proper calibration between the imaging sensors.

4.2.2. Software Platform and Architecture

We planned to develop the proposed system on the PC platform running Microsoft Windows 7 operating system, using the Unity 3D v4.5 game engine. The main visualization software will be developed as a Unity 3D project, while its integration with the gesture interaction module will be held through developing a plug-in for Unity. The plug-in will be developed using Visual Studio C++ 2010 with Microsoft Kinect SDK v1.7. As of the depth sensing camera, we will use Kinect for Windows (v1) which provides 1280 x 720 resolutions of RGB image stream, as well as 640 x 480 resolutions of depth image stream. The system will use an additional HD web cam in case needing high resolution live video to be used as a background for virtual mirror style visualization.

Based on the software development platform described above, we designed the software architecture design for implementation. Figure 4.3 shows the overview of the software architecture design.

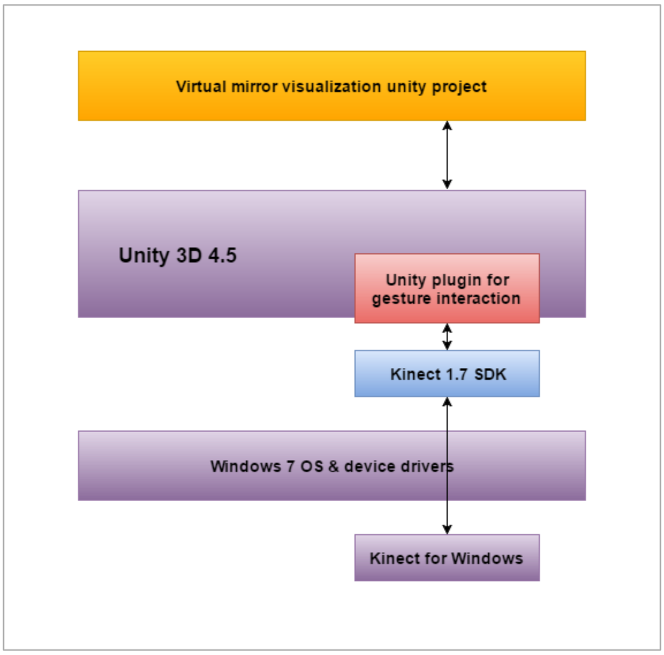

Figure 4.3: Software architecture

The research project mainly developed two components in the software architecture diagram: the virtual mirror visualization unity project and the unity plug-in for gesture interaction. The Unity 3D project implements virtual mirror visualization and content for interaction, while the Unity plug-in allows the Unity 3D project to access Kinect sensor data based on the APIs provided by the Kinect SDK.

4.3. Implementation

As a proof of concept, we implemented a sample prototype application that shows a hat on the user's head in a virtual mirror like visualization. For testing the interaction method we added a button that changes the color of the hat when selected.

We implemented the proposed UI agent guided gesture method together with other alternatives used as conventional methods to compare them. Figure 4.4 shows the UI that provides no feedback until the Color button is triggered after a certain amount of delay, while Figure 4.5 shows the UI that demonstrates this delay with a simple timer represented by the button getting filled up in different color.

Figure 4.4 Prototype implementation – no feedback

Figure 4.5 Prototype implementation – box filling timer

Figure 4.6 shows a delayed pointer UI where the simple circular pointer follows the user's hand with certain amount of delay, while Figure 4.7 shows our proposed method using a software UI agent following the user's hand. The butterfly has an animated sequence of wing flapping to add more life to the character.

Figure 4.6 Prototype implementation – pointer following hand with delay

Figure 4.7 Prototype implementation – UI agent following hand with delay

5. Focus Group User Study

We conducted a focus group user study to collect feedback from potential users to discuss and share ideas on how gesture interaction with public displays can be more compelling. The study also allowed discussion of the usage of UI agents such as virtual characters to improve user experience.

5.1. Procedure

The focus group included answering questionnaire, group discussion, and brief demonstration of prototype interface. The entire process for the focus group took about an hour and a half. At the beginning of the focus group, we asked the participants to fill out the questionnaire to collect demographic information and the participant's background regarding their use of public information systems, gesture interfaces, and their preference on virtual characters. The questionnaire used can be found in Appendix A of this report.

The participants were asked to discuss various topics around four themes: public information systems, gesture interfaces, UI agents and virtual characters. Also, how they can be designed for gesture interfaces to improve the experience of using public information systems. The demonstration was held after discussing on the first three themes of topics where we showed the participants, the prototype user interface described in the previous chapter. During discussions, we also asked participants to vote on some of the topics to collect quantitative measures of the participants' opinion.

5.2. Results

Thirteen graduate students participated in the focus group. They were 29.6 years old on average (*Std. Dev.* = 4.4) and three of them were female. In the following subsections we summarize the results of the questionnaire responses, and the verbal feedback collected around the topics under the four themes for group discussion.

5.2.1. Questionnaire responses

Regarding the results from the questionnaire, when asked how often they use public informati on systems, most of them (11 participants, 85%) answered few times a year, while the other t wo (15%) participants using more frequently, few times a week. The most popular case of usi ng public information system was for accessing guide information for tourist or visitors in var ious places (airport, bus and train stations, shopping mall, etc.) mentioned by 9 participants (6 9%). This followed by museum installations (7 participants), library catalog (5), and other self-check-in/out kiosks at stores or airport (see Figure 5.1). When asked to rate on an 11-point Likert scale (0: totally disagree ~ 10: totally agree) if they actively try to use interactive public displays, three of the participants answered lower than the neural rating (5) while the r est rated 7 or above (see Figure 5.2), although the one sample Wilcoxon Signed Rank test did not show significant difference from the neutral rating value (W = 55, p = .0574, *Median* = 7, *Inter-Quartile Range* = [7-8]).

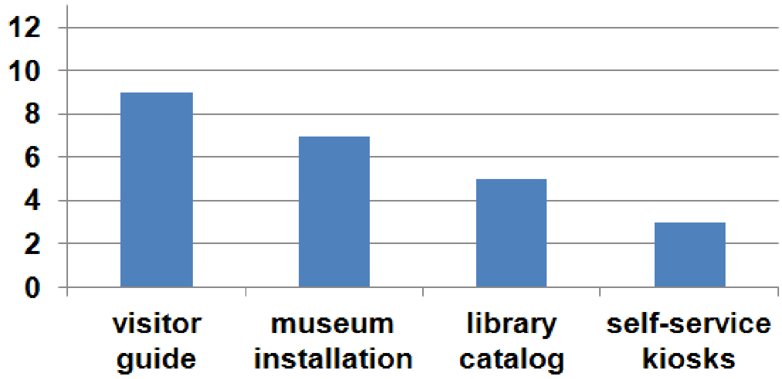

Figure 5.1 Type of public information system used

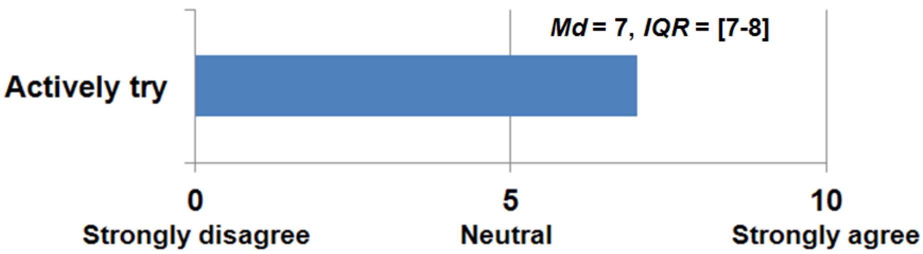

Figure 5.2 How much actively try public information systems

The participants appeared to be using gesture interfaces not much frequently, as nine of the participants (69%) were using it few times a year or less, while four (31%) using it few times a month (see Figure 5.3). When asked what type of gesture interfaces they have used (see Figure 5.4), the Microsoft Xbox Kinect was the post popular being mentioned by 9 participants (69%), while few participants mentioned Wii (3 participants, 23%) or Leap Motion (2 participants, 15%). When asked to rate on an 11-point Likert scale if they are good at figuring out how to use gesture interfaces, the participants thought they were fairly confident as all of the participants rated above the neutral rating (5). The median rating value was 8 (IQR = [7-

8]), and the one sample Wilcoxon Signed Rank test showed it was significantly different from the neutral rating (W = 91, p = .0016).

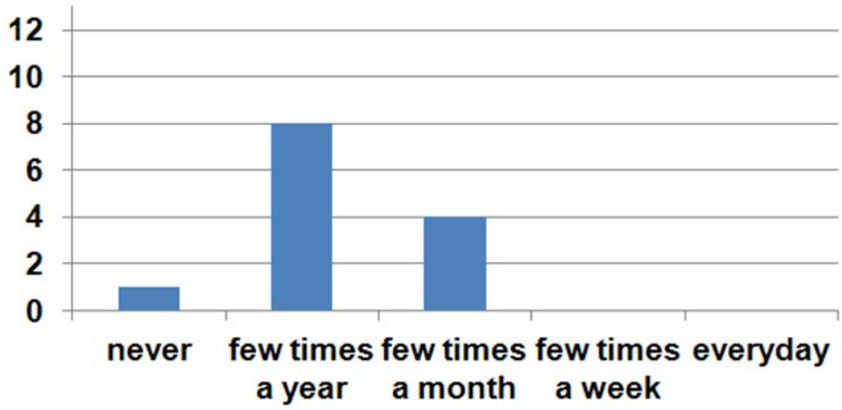

Figure 5.3 How much frequently using gesture interfaces

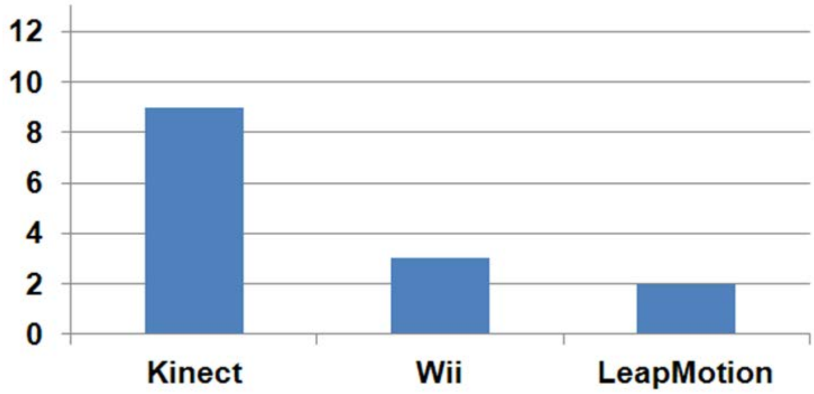

Figure 5.4 Type of gesture interface used

Most of the participants (10 participants, 70%) stated they love characters by rating 7 or higher on an 11-point Likert scale, while other three rated equal or lower to the neutral value 5 (see Figure 5.5). The rating was in overall significantly higher than the neutral value based on a Wilcoxon Signed Rank test (W = 68, p = .008, Md = 8, IQR = [6-10]). When asked what are their favorite characters, characters from animations or cartoons were the most popular mentioned by 5 participants, while those from live action movies (2 participants) and games (2 participants) followed (see Figure 5.6). Most of the participants (9) agreed on that they are affected by products with their favorite characters by rating 7 or above on an 11-point Likert scale, while the other four rated 5 or below, although in overall it was not significantly different from rating on the neutral value based on the one sample Wilcoxon Signed Rank test (W = 47, p = .0688, Md = 7, IQR = [5-7], see Figure 5.7).

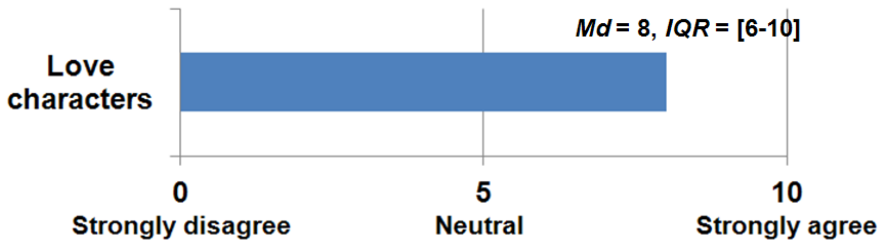

Figure 5.5 How much loving characters

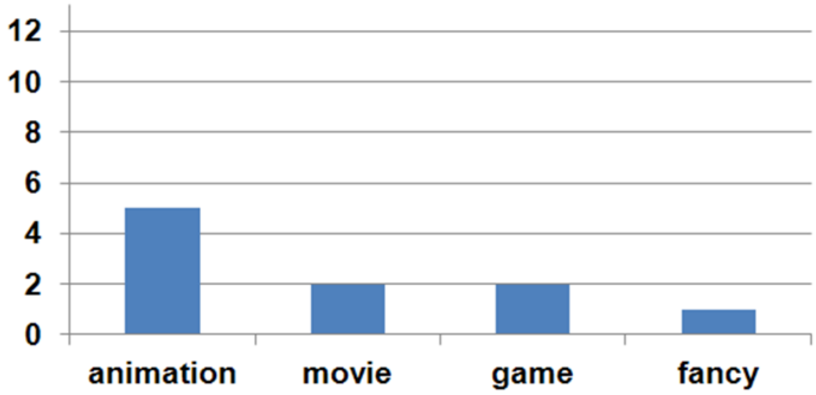

Figure 5.6 Type of favorite characters

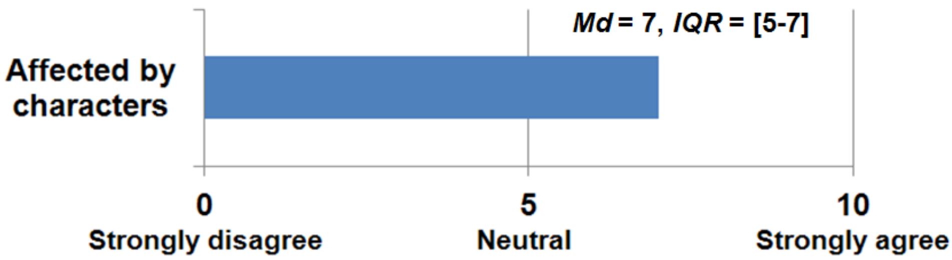

Figure 5.7 How much affected by characters

5.2.2. Interactive public information systems

The discussion started with sharing previous experiences of using interactive public information displays, how good or bad it was. Participants shared various cases including touch screen based interactive kiosks at shopping centers for browsing shop directories and finding was to get there, a self-service checkout machine for paying in a store, and a ticketing machine at a train station. Besides these commercial uses of kiosks for automating transactions, participants also shared other cases that are more focusing on interactivity itself to provide compelling experiences to the public. Some of the examples shared by the participants included an interactive installation at an airport where people can post messages on a wall display using gesture interfaces, watching a virtual globe on an immersive interactive kiosk at a library, and interactive floor projections with advertisements with simple games.

Participants mostly agreed on there are many interactive public information kiosks becoming more widely accessible, yet there are still some cases where the system is not simple enough for general public to walk up and use it easily. One of the participants emphasized on this saying, “it should be simple, and have as few steps as possible” regarding kiosks for automated commercial transactions. Another participant shared his experience with trying to buy a train ticket on a kiosk in a foreign country, and failing to find proper language option and ending up not being able to figure out how the system works. One of the participants suggested the interactive public information systems should have feedback collection features so that those who are having problem with using them should be able to give feedback on how to improve them.

When asked how likely they would be willing to try an interactive public displays besides those that are necessary to use for commercial transactions, participants responded it will be depending on how much interesting is the content shown on the display. One of the participants mentioned that he would not try if it’s a clear advertisement material that simply tries to sell something. Yet if the content is interesting there might be higher chance of trying it. Another participant mentioned it might be depending on the context including being alone or together with friends, not willing to be stared at by others when using it alone but with friends could try it out for fun. Besides personal interest on the content, few participants raised the point of looking abnormal or unusual things also catches people’s interest out of its novelty, saying “people like to try if it’s beyond common sense.”

People being interested in unusual things raised a discussion on using novel interaction methods. While many of the interactive kiosks use touch screens these days, still huge touch tables that many people can interact with simultaneously or user’s mobile devices used in combination with touch tables were considered as interesting options.

Gesture based interaction was also considered as a novel way to interact with public displays by the participants, despite of its problem of not being accurate or efficient enough for certain tasks, such as giving text input. Voice recognition was discussed as an alternative, with one of the participants saying, “Gesture is tiring, just say something.” Yet some other participants raised the point of voice recognition also having limitations with accuracy especially when used public spaces. One of the participants suggested public information systems using novel interaction methods should provide alternative methods to fall back when the novel method fails to work.

5.2.3. Gesture interfaces

Although it became more popular nowadays, gesture or motion based games were still considered as novel and new with only three participants (23%) saying they play such games. Based on binomial test, having 3 or less participants not playing is significantly less than having an equal chance of playing or not (p = .0461). The reasons participants pointed out why they are not widely adopted yet included being expensive, requiring large spaces to setup, and not many content (games) being available.

Besides not being widely available, other problems of gesture interfaces mentioned by the participants included not being accurate enough to do serious tasks, tiring with causing physical fatigue when using for long period of time, and social aspects of looking weird with making gestures in the air when playing alone. The gesture based games were considered to be more acceptable when playing together with friends in a group. Another problem of gesture interaction discussed was not being intuitive as much as expected. When asked how many of the participants had troubles of not knowing what to do using a gesture interface, almost half of the participants (6 participants, 46%) answered they experienced problems such as not being sure if they needed to do a pressing gesture or simply hold their hand at an icon to select it.

Despite of its limitations, gesture based interface was stilled considered as one of the fun and novel interaction methods for public interactive installations by providing means of physical interaction with users. Floor projected interactive advertisements were discussed as one of the fun public interactive system that uses gesture recognition technology, while other examples that uses simpler technology yet providing compelling physical interaction were discussed. For example, floor piano, or stair cases responding to users' steps were identified as one of the traditional success cases of providing compelling physical interaction with simple technology.

When asked if they would be more willing to try using a public information system when using a gesture interface, most of the participants (10 participants, 77%) answered they would (binomial test compared to equal chance of choice: p = .0461) given that the content shown looks interesting or fun, rather than for serious applications.

5.2.4. UI agents and virtual characters

The third theme of discussion topic was around virtual characters and UI agents. Only two of the participant said they or not much interested in virtual characters while the rest of the participants expressed their general interest. About 70% of the participants said they have favorite characters. Some of the participants mentioned that they consider characters are much more appealing to children.

Despite of overall interest in virtual characters, most of the participants stated they did not like 'Clippy' and other UI Agents in Microsoft Office that were widely available few years ago. Only two participants state they enjoyed having them (binomial test compared to equal chance of liking: $p = .0112$).

The main reason raised by the participants on why this famous UI agent failed was not meeting up with users' expectation. One of the participants mentioned, "(it was) occupying more screen than (how much) it is useful," while another participant said, "it isn't much helpful as expected."

Based on what Clippy was able to do, participants thought it was only for very novice users such as first time users, but not for advanced users. Participants said it was very annoying when it actively suggested giving help even when the user didn't need any. One of the participants said it was distracting as it was unnecessarily moving when she needed to focus on working. Another participant said it would have only used for initial briefing or tutorial but not for every time.

Compared to the famous failure of applying a UI agent, some of the participants shared some positive cases, such as Jess on the Jetstar[6] website. Jess is a virtual assistant represented by a simple image without any animation. When a user clicks on Jess it opens up a chat box where people can type in questions to ask regarding flight schedules or booking. Jess simply shows search results over the website, yet it adds few phrases that make the experience compelling enough as if the user is chatting with a virtual agent. Participants mentioned the simple, non-distracting, and helpful yet not overselling what it can do is building the positive experience when interacting with Jess.

Despite the limitations and problems discussed regarding Clippy, participants agreed on that

[6] http://www.jetstar.co.nz

some of those problems might not be an issue when applying UI agents to public information systems, as they are mostly used by novice users if not the first time users. Moreover, their nature of distracting users by catching attention with animation were considered as it could be even useful for attracting potential users to try using the system.

When asked if they would more likely try a public information system when it shows a virtual character on it, 10 participants (77%) answered they would (binomial test compared to equal chance of choice: $p = .0461$) given that the character is nicely designed and interesting.

5.2.5. Using UI agents with gesture interfaces on public info systems

After showing a demonstration of the prototype system described in the previous chapter, we collected feedback from the participants on using UI agents in combination with gesture interaction.

In overall, participants saw the benefit of having UI agents for improving gesture interaction yet given that the virtual character used is relevant to the content and application. One of the participants pointed out that if the character used is not preferable by itself (e.g. poorly designed or socially unacceptable) then the user experience could get worse.

The clearest advantage of having animated virtual character responding to user’s hand motion participants saw was helping the users to clearly be aware of the system tracking user’s motion. As the virtual character was constantly following the user’s hand, participants naturally understood that the systems tracks and responds to hand motion, so that the user has to use his or her hands to direct the virtual character where to go, even without showing a pointer on the tracked user’s hand.

Regarding selecting a button, while the virtual character following the user’s hand with certain amount of delay did suggested the user to hold his/her hand at the button until the character arrives, the alternative design with box filling up was perceived as giving more prompt feedback as the user’s hand hovers on the button. This suggests we could improve the implementation using UI agent by highlighting the button when the user’s hand is on it.

Some participants suggested the amount of time delay for confirming the selection needs to be carefully decided. The current implementation was using 2 seconds delay which could be perceived as too long when using the system for longer period. Another feedback to improve the prototype system was on designing the button to have a shape relevant to the virtual character. With current simple square button, participants thought the relationship between the butterfly and the button was weak.

In overall, the virtual character was perceived as a good way for attracting and grabbing attention from the user, and encouraging the user to interact with motions. The proposed method of using gesture interaction was perceived as to be more appropriate and useful for light applications such as entertainment or simple games, but not relevant for serious applications that needs productivity.

By improving the proposed interaction method based on the feedback collected in the focus group user study, we believe guiding users on public information systems could become one of the representative tasks that UI agents are helpful and can perform well.

6. Prototype System v2

Based on the results of the focus group in chapter 5, the project continued on another iteration of the design and implementation process for building a second prototype system based on the use case scenarios described in section 3.2. This chapter describes the details of the design and implementation of the second prototype system.

6.1. Overview

As the target application of the prototype system, we chose a simple quiz application where users can give simple answers (e.g. true/false, or yes/no) to questions around certain domain knowledge. This was chosen for focusing on the button selection tasks. The users were to answer by selecting a button of an answer that they thought it was correct. One of the requirements was to use a virtual character that will be used in the sponsor's main project. Figure 6.1 shows an image of the virtual character used, a robot goose. Based on the virtual character to be used in the prototype system, we chose the theme of the quizzes around the topic of robots and birds.

Figure 6.1 A robot goose virtual character used for the prototype system

According to the chosen virtual character and initial feedback from the users at the focus group, we designed and implemented the prototype system following the use case scenario design described earlier in section 4.2. While initially there were four scenes designed in the use case scenario, we chose to focus on the first three as those were identified as the main cases that would be widely required in many different types of public information systems, and also as they are the main interest of the sponsor's main project.

Based on the three cases, we defined the interaction flow of the prototype system into three stages: idle, engagement, and main application. As illustrated in Figure 6.2, the prototype system initial starts from the idle stage where it waits for a user to arrive and start using it. Once a user is identified, it transitions onto the engagement stage where the system helps the user to get ready for using the main application. After the user finishes using the main application and leaves, the system moves back into the idle stage where it waits for the next user.

Figure 6.2 Interactive stages in the prototype system

6.2. Idle Stage

The idle stage is where a public information display is waiting for a user to come and engage with it for further use of the system. In addition to showing public information (such as advertisement) that could catch interest of crowd passing by the public information display, the proposed prototype system is designed to actively grab attention of people passing by through using interactive UI agents. While the initial design had two options, there was some feedback that concerned of the virtual character being hardly noticeable when they appear at the bottom of the display. Based on this feedback, we decided to design the idle stage for the prototype system to be a virtual robot goose character flying and following a potential user's face (i.e. the closest person in front of the public information display). In order to highlight the user's face, a hole was visualized on top of the idle screen with an advertisement image, where a potential user's face becomes visible as if it is reflected on a mirror. Figure 6.3 shows screenshots of the prototype system at the idle stage. The system tracks the user's head and neck joint positions to decide the size of the hole to be big enough to encompass the user's whole face. However, the system maintains the size of the hole to be smaller than a certain size so that if a person gets too close to the screen they will start seeing only portion of their faces, suggesting to keep a certain distance away from the screen.

Figure 6.3 Prototype system at the idle stage

As a person passing by catches an interest, he or she would stand and face towards the public information display screen, which would be a trigger for moving onto the engagement stage. To detect a potential user's interest, the prototype system tracks not only the face position of the user but also his or her shoulders to determine if the user's body is facing towards the screen or not. To determine whether the user's body is facing towards the screen or not, we check if the vector between the user's shoulders are tangent to the screen surface. Given that the screen surface is on the XY plane, in order to check if the user's body is facing towards the screen the system simply checks the following equation,

$$|\mathbf{vs}_z| < \textit{threshold}$$

where $\mathbf{vs}_z$ is the Z component of a vector between the user's shoulders. For our implementation, we use 15 centimeters as the *threshold* value.

In order to ensure the user is indeed expressing his or her interest, the prototype system also checks the duration of user facing towards the screen. The prototype system checks if the user is facing towards the screen for more than three seconds, then it moves on to the engagement stage. During these three seconds, the system enlarges the hole, where the video image of the user's environment is shown, overtime to give feedback to the user on if the user is recognizing his or her expression of interest (see Figure 6.4).

Figure 6.4 Detecting a potential user's interest and transitioning to the engagement stage

6.3. Engagement Stage

Once a user expresses his or her interest through standing in front of the prototype system and facing towards the screen, the system should guide the user to stand at the ideal place for further interaction with it. The prototype system uses both graphical symbols and the UI agent to guide the user to the place to stand. Figure 6.5 shows the graphical symbols, a red outline of a footprint on the floor and a yellow arrow animated to point at the footprint. In addition to these graphical symbols, the prototype system also lets the UI agent to come at the footprint to draw user's attention.

To check if the user is standing at the right position, the prototype system measures the distances between the footprint symbol and the two (left and right) ankle joints of the user. If both of the ankles are within certain distance (e.g. 25 centimeters in our implementation) from the footprint symbol, the prototype system considers the user is standing at the right place. Similar to the transition from the idle stage to the engagement stage, the system also checks how long the user has been staying at the right spot to prevent accidentally transitioning into the next stage.

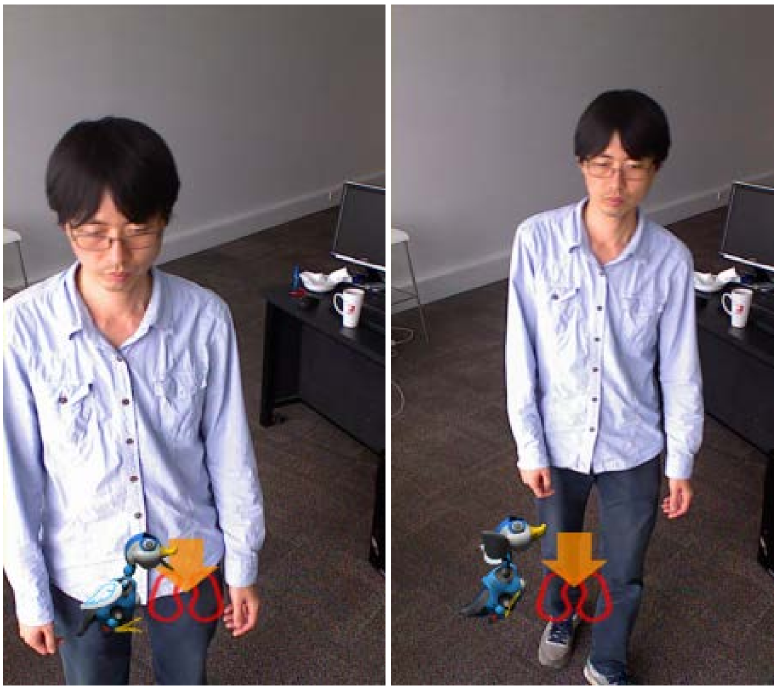

Figure 6.5 Guiding a user to stand at the ideal place in the engagement stage

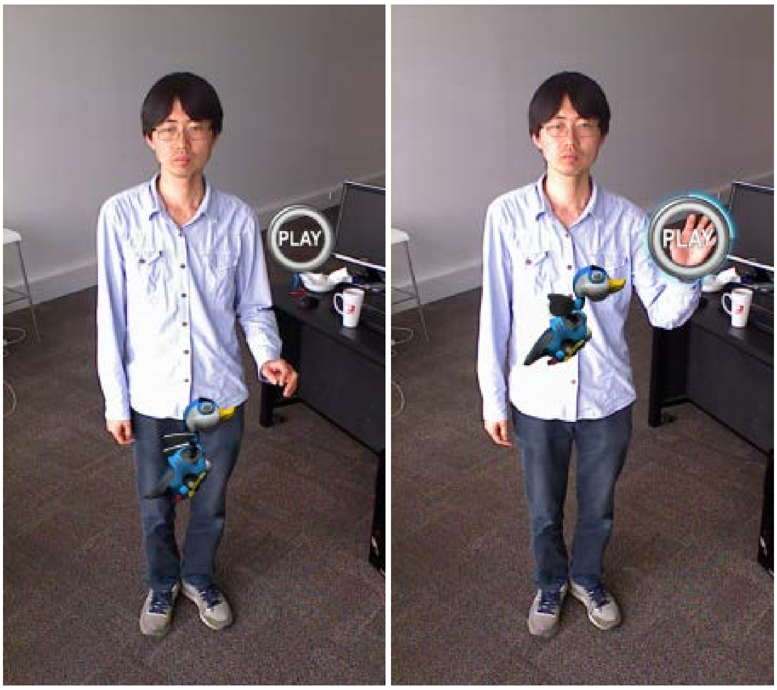

Figure 6.6 Start button to begin the main interactive application

6.4. Prototype

The main application of the prototype system is a simple quiz to be answered by choosing a button with the right answer. The theme of the quiz was chosen as birds and robots to reflect the virtual character used as a UI agent. The main application part starts with a screen with a 'Start' button to begin the quiz (see Figure 6.6). At this screen, the user not only can have a moment to be ready before starting the quiz, but also learns the very basic but essential interaction which is selecting a button.

Selecting a button is done by hovering user's hand over a button and holding it in position for a moment (2 seconds in our implementation). As one of the user's hands hovers over a button, the button gets highlighted and scales up a bit as an indication of hovering (second from the left of Figure 6.7). The timer indicator around the button fills up in clockwise direction for the amount of time user has to hold his or her hand in place. Once the allocated time has passed, the button glows brighter to indicate its selection.

Figure 6.7 Button animation for selection

Once the user selects the Start button, a brief instruction on how to answer the quiz (as on the left of Figure 6.8) is shown on the screen of the prototype system. After the instruction, the quiz starts and the system shows a quiz questions with a set of buttons (e.g. Yes/No) to choose an answer from (see right of Figure 6.8)

Once the user selects an answer, the answer buttons disappear and the screen shows whether the chosen answer is correct or not (see Figure 6.9). After a few seconds, the prototype system shows the next question for the quiz and repeats until all of the quiz questions are answered. And finally shows a 'Thank you!' message, and the system turns back to the idle stage after few seconds.

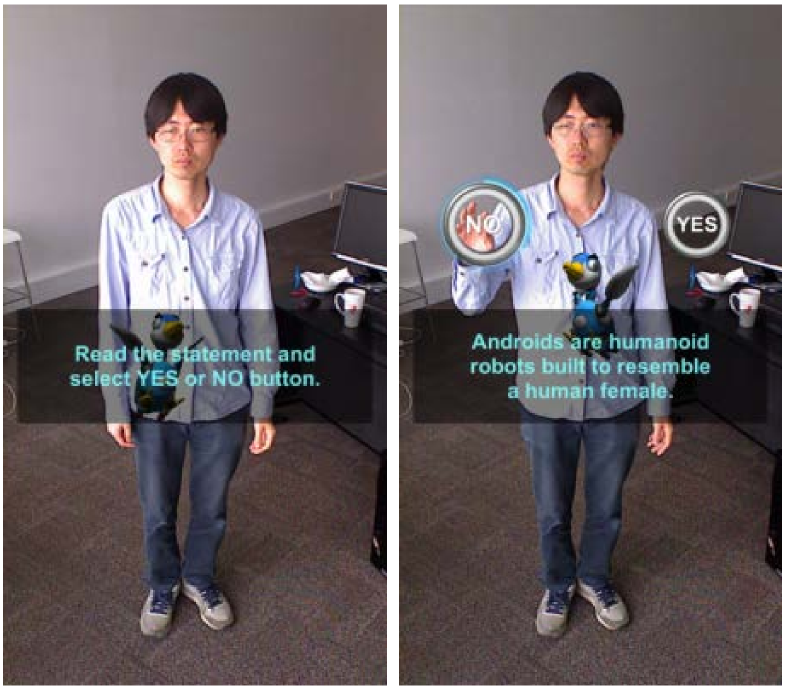

Figure 6.8 Instruction and quiz screens

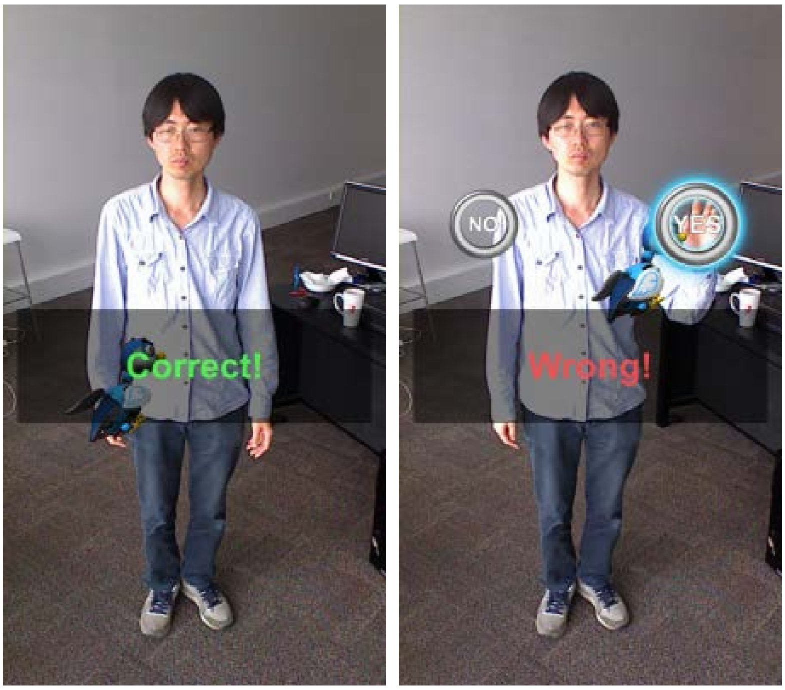

Figure 6.9 The screen showing whether the answer is correct or wrong

6.5. UI Agent Behavior

The main role of the UI Agent (i.e. the robot goose virtual character) in the prototype system is to guide the user for interacting with the system. This includes following a potential user's face in the idle stage, guiding the user's attention to the place to stand in the engagement stage, and giving hints on how to select a button using gestures. In overall, the basic behavior of the

UI agent required in the prototype system is following a target position or object, and being at assigned places at a certain time according to the application scenario.

When there is no gesture input expected from the user, for example when instruction or a message is shown on the screen, the system lets the UI agent to wander around without following any target. On the other hand, when the system needs certain input from the user, the UI agent follows certain target objects or stays at a certain position to draw attention of the user and guide the interaction.

The behavior of the UI agent following a target or being present at the target at a certain time is implemented based on the model described in section 5.1. In our prototype system implementation, temporal constraints (when the UI agents arrives at the target) can be defined as either the amount of time left (as in section 5.1) or the target time of arrival in the simulation time. The amount of time left can be calculated by the following equation,

$$t_l = t_a - t_c$$

where t_l is the amount of time left, t_a is the target time of arrival, and t_c is the current time in simulation time.

In the idle stage where the UI agent follows the user's face, the amount of time left for the UI agent to be at the target position is set as a constant value (e.g. 1 second in our implementation), and the target position is updated by the tracked position of the user's head joint. In this way, the UI agent behaves as if it is trying to catch up the user's movement faster when it is further away, while slowing down when it reaches the target (see Figure 6.10).

In the engagement stage, the target position of the UI agent is set to the footprint symbol with the target time of arrival set to 1 second past from the current simulation time, which is:

$$t_a = t_c + 1$$

This ensures the UI agent to move to the footprint within a second to draw the attention of the user to the footprint symbol (See Figure 6.5).

When the user has to select a button, the UI agent's target object is set to the tracked user's hand joint with the amount of time left set to a constant value (e.g. 1 second), as a result following the user's hand. This behavior suggests the user that the UI agent is interested in the user’s hand motion. When the system allows the user to use both hands to interact with, the system can decide which hand the UI agent should follow based on heuristic rules. In our prototype system, the hand that is placed higher is given priority for the UI agent to follow (see the two images on the left of Figure 6.10), and in addition the right hand (which is the dominant hand of most people in general) is given higher priority over the left hand.

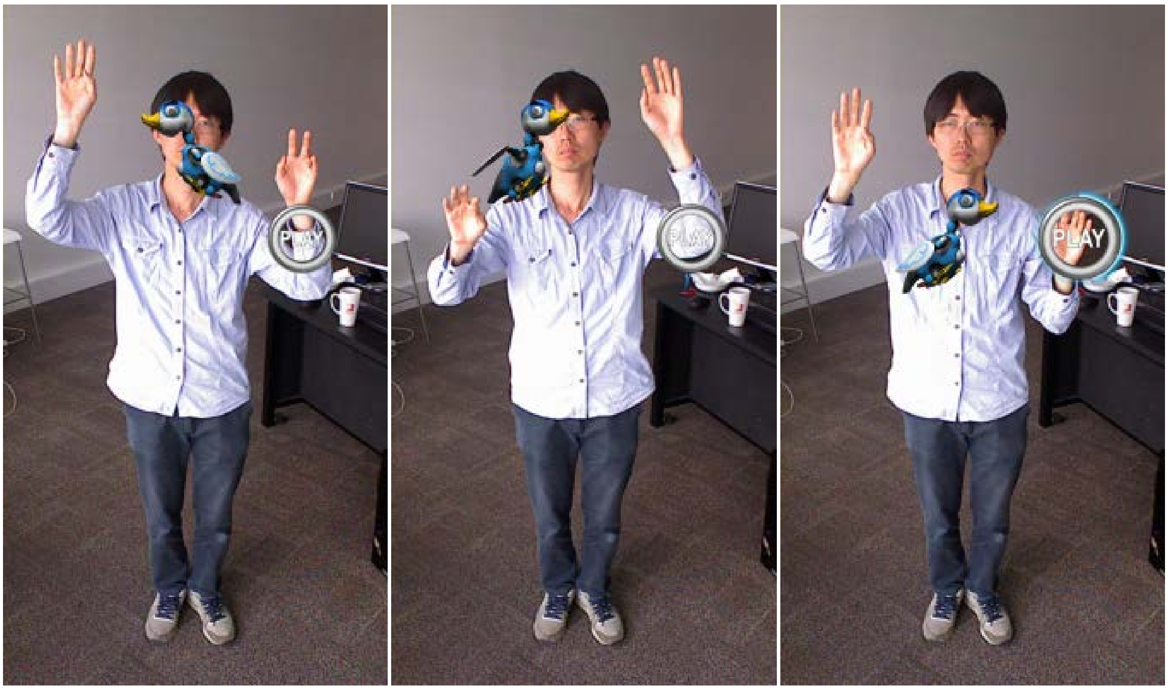

Figure 6.10 UI agent following the hand that is raised higher, or the button if hovered on

When the user’s hand hovers over a button, the button becomes the target object of the UI agent (see the right image of Figure 7.10), and the target time of arrival is set to the point when the button will get selected after holding the hand on the button for a while. In other words, if the amount of time needed to hold the hand on a button to select is t_h and the current time is t_c, then the target time of arrival of the UI agent t_a can be calculated as:

$$t_a = t_c + t_h$$

In this way, the button will get selected as the UI agent reaches the button, creating an illusion of the UI agent selecting the button. After few pilot trials, we learnt from users’ feedback that it is better to have the UI agent arrive a little bit earlier than when the button selection happens, and also to give certain feedback (e.g. an animation of the UI agent trying to select the button) for helping users to perceive as if the UI agent is selecting the button. In our prototype system, we applied this by making the UI agent arrive about a half a second earlier than when the button is selected, and making animation of waving their legs and making quack sounds as arriving at the button (see Figure 6.11).

Figure 6.11 UI agent feedback when reaching a button to select

7. User Evaluation of the Prototype v1

To evaluate the proposed user interface design, we conducted an experimental user study with the prototype system implemented. The goal of the user study was to evaluate the effects of different visual cues on user's experience while interacting with the virtual mirror interface using gestures. The user study design was reviewed and approved by the Human Ethics Committee at the host institution according to Low Risk process. This chapter describes the details of the experimental design, report on the results, and discuss on the implications of the findings from the user study.

7.1. Setup

The experiment was held in a laboratory space with the prototype system installed and a desk and chairs. The user was asked to stand in front of the prototype system to do the experimental tasks and sit on the chair for receiving instructions, answering questionnaires, and debriefing. Figure 7.1 shows the environment of the room we conducted the experiment. The prototype system setup included a 55 inch HD TV screen mounted in portrait orientation on a mobile stand, with a Microsoft Kinect motion sensor mounted at the top of the TV screen. A PC with Intel Xeon E3-1240 3.4GHz CPU, 16GB of Main Memory, and NVidia GeForce GTX 750 GPU operated with Microsoft Windows 7 operating system was used as the main computer to run the prototype software.

Figure 7.1 Experimental setup

7.2. Experimental Design

The user study was designed as a formal experiment, with the main independent variable as the design of visual cues used with the prototype system. We compared three types of visual cues: None, Graphical Symbol, and UI Agent. The None condition (N) provided no (or minimum) visual cue as the participant were to figure out what to do or how to interact with prototype system, while in the Graphical Symbol condition (G) the prototype system showed visual cues of graphical symbols on guiding the user what to do or how to interact with. In the UI Agent condition (A), the proposed method of using UI agents as visual cues was applied to guide the user with interaction. Table 7.1 summarizes the difference between the conditions in more details. (For more detailed design of the visual cues, please read section 6.2 to 6.5)

The task was to use and interact with the prototype system based on what is shown on the screen. While we used within-subject design for investigating the Main Application stage (answering quiz questions on the topic of birds and robots), we chose between-subject design for the Idle and Engagement stages as the main focus of the investigation is around intuitiveness which requires participants to be new to the system (i.e. never used it before).

Table 7.1 Visual cues in the experimental conditions

Stage	None	Graphical Symbol	UI Agent
Idle	- Full screen image background	- Full screen image background - A hole of video stream highlighting user's face	- Full screen image background - A hole of video stream highlighting user's face - UI agent following user's face
Engagement	- Red footprint outline	- Red footprint outline -Yellow animated arrow pointing at the footprint	- Red footprint outline - UI agent staying at the footprint
Main Application	- Button highlighted on hover - Button glows brighter on selection	- Button highlighted on hover - Button glows brighter on selection - Timer indicator	- Button highlighted on hover - Button glows brighter on selection - UI agent following user's hand and selecting the buttons

7.2.1. Procedure

The user experiment followed the procedure summarized in Table 7.2. The researcher followed the script (see Appendix B) for dealing with the participants in consistent manner. An experimental session started with welcoming a participant to the experimental environment. The participant was then asked to sit and read the information sheet (see Appendix C), and sign the consent form (see Appendix D). After signing the consent form, the participant was asked to fill in the pre-experiment questionnaire (see Appendix E) which asked demographic information and background of the participant, before continuing on to the experimental trials.

The first section of the experimental trials was for investigating on how intuitive is the system for the first time users. Participants were asked to stand in front of the prototype system and interact with it only based on what is shown on the screen but without any further instruction from the researcher. In this trial, the system started from the Idle stage and continued on up to selecting the Start button in the Main Application stage. As this section was in between-subject design, each participant experienced only one of the experimental conditions. After finishing the task, participants were asked to sit at the table and answer the per-trial questionnaire (see Appendix F). And then a brief interview followed which asked about their experience of interacting with the system for the first time.

Table 7.2 User Evaluation Study Procedure

Duration	Procedure
5 min.	- Welcome - Informed consent - Pre-experiment questionnaire
Section 1 (between-subject design)	
5 min.	- Initial trial of the prototype system in one condition - Per-trial questionnaire - Interview
Section 2 (within-subject design)	
15 min.	Repeat for three conditions: - Trial of the prototype system in a condition - Per-trial questionnaire
5 min.	Post-experiment questionnaire
5 min.	Interview & debriefing

The second section of the experimental trials was in within-subject design. The participants had three trials of using the prototype system in different conditions. The order of conditions was counterbalanced using Balanced Latin Square design. In each trial, the system started from the Engagement stage where it indicates the participant where to stand. After standing at the position, the participant continued on to the other stages following the information provided on the screen. Ten quiz questions on the topic of birds and robots were asked in each trial where the participants had to answer by selecting the Yes or No buttons (see section 7.4 for more details). Different set of quiz questions were used in each trial. After answering ten questions the system showed a 'Thank you!' message for concluding the trial. After each trial, participants were asked to answer the per-trial questionnaire (see Appendix F). After finishing all three conditions, participants were asked to answer the post-experiment questionnaire (see Appendix G). Finally, the participants were interviewed briefly before concluding the experiment.

7.2.2. Measurements

The main method of measurement we used was using questionnaires. For each trial we used System Usability Scale (SUS) [40] [41] and O'Brien's Engagement [42] questionnaires to measure the usability of the prototype system and the user's level of engagement. The SUS questionnaire results were aggregated into the range of 0~100 and the Engagement questionnaire results were aggregated into the range of 1~5. A semi-structured interview was conducted in each section of trials, asking various questions such as how easy it was to know how to interact with the system, any perceived difference in the time delay for selection, and how did the UI agent character affected the user experience.

In addition to subjective measures on usability and user experience, we also collected objective measures through observation, such as number of time the participant tried wrong gestures for selecting a button and the type of gestures tried, which hand was used for interacting with the system, and if the participant had troubles in using the system. Besides video recording the experimental trials, we also collected system logs of events such as the beginning and the end of each stage, and when a button is hovered on/off or selected.

7.3. Participants

We recruited 21 participants through advertising on the university campus and online community websites. The participants were between 19 to 34 years old (*Mean* = 26.2, *Standard Deviation* = 4.36, *Median* = 25) and 10 of them were female (52.4%). Most of the participants used right hand as their dominant hand (*N* = 18, 85.7%) while two participants answered they use both hands, and one participant answered the left hand.

In the pre-experiment questionnaire (see Appendix E), when asked how frequently they use public information systems, more than half of the participants (*N* = 12, 57.1%) answered a few times a year, but there was none who answered ‘every day’.

The participants were not much familiar with gesture based interfaces, as when asked if they have used gesture interface before, most of the participants (19, 90.5%) answered less than few times a year or not at all. When asked if they have played Microsoft XBOX Kinect motion games before, 7 of the participants (33.3%) answered they have not played at all, while more than half of the participants (*N* = 11, 52.4%) answered few times a year and only 3 of the participants (*N* = 3, 14.3%) answered few times a week. The response was similar when asked if they have played Nintendo Wii or Sony MOVE, where 9 of the participants (42.9%) answered not at all, and 12 of the participants (57.1%) answered few times a year.

When asked if they have used Augmented Reality (AR) app or interface before, 5 of them (23.8%) answered they were not aware of what AR is, while the rest (*N* = 16, 72.8%) had used at least few times a year (the mode was ‘few times a month’ as 7 participants has chosen this answer).

There were three 7-point Likert item questions (1: Strongly disagree ~ 7: Strong agree; 4: neutral) in the pre-experiment questionnaire, which showed participants saw themselves moderately using gestures in everyday life (*Median* = 5, *Inter-Quartile Range* = [3-5]), positive about characters (*Median* = 6, *IQR* = [5-7]), yet they had diverging opinions on if their favorite characters affected their experience with a product (*Median* = 4.5, *IQR* = [2-6]).

7.4. Results

Here we report the results of the per-trial and post-experiment questionnaires, and also summarize the findings from the interviews and observations. The per-trial questionnaires (System Usability Scale and Engagement questionnaires) results are reported separate sections. The first section (section 1) of trials focused on the first time use of the system. The first section was in between-subject design, while the second section (section 2) was in within-subject design focusing on the main application. See section 8.2 for more details of the experimental design. All inferential statistics were tested with an alpha level of 0.05, unless it is stated otherwise.8.4.1 System Usability Scale

For section 1 trials, all three conditions showed slightly above the average (70) usability based on the System Usability Scale (SUS) results (see Figure 7.2). The SUS average score for the None condition (N) was 71.8 (*Md* = 70), while the Graphical Symbol condition (G) scored 70 on average (*Md* = 72.5), and the UI Agent condition (A) scored 71.4 (*Md* = 75). No significant difference was found between the conditions based on Kruskal-Wallis test ($H = 0.067$, $p = .967$).

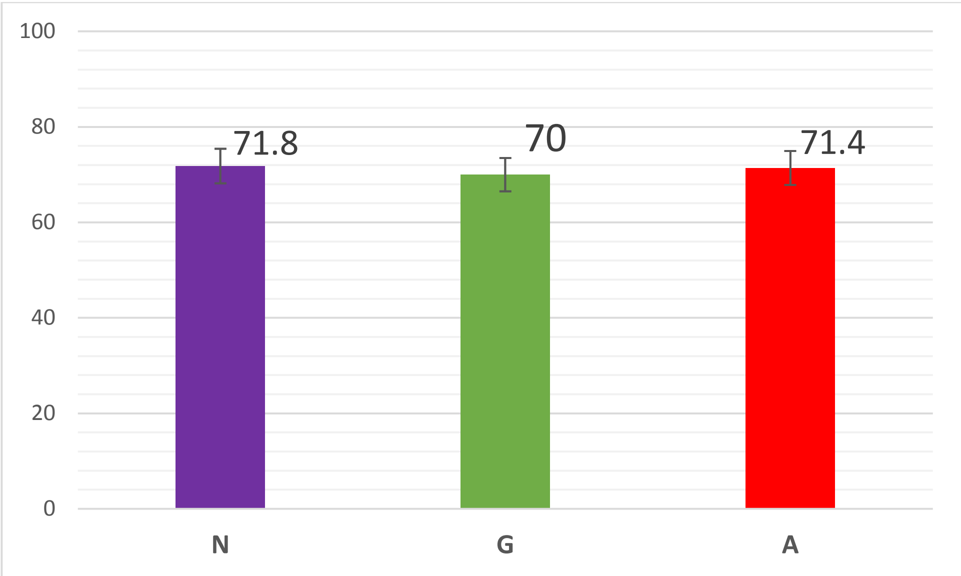

Figure 7.2 Section 1 results of the System Usability Scale

The results from the section 2 trials were also similar with all the conditions showing slightly above the average SUS score. The average SUS scores of each condition were 76.2 (*Md* = 80) for the None condition, 82 (*Md* = 82.5) for the Graphical Symbol condition, and 75 (*Md* = 77.5) for the UI agent condition. There was no significant difference found between the conditions based on Friedman test ($\chi^2(2) = 3.552$, $p = .169$).

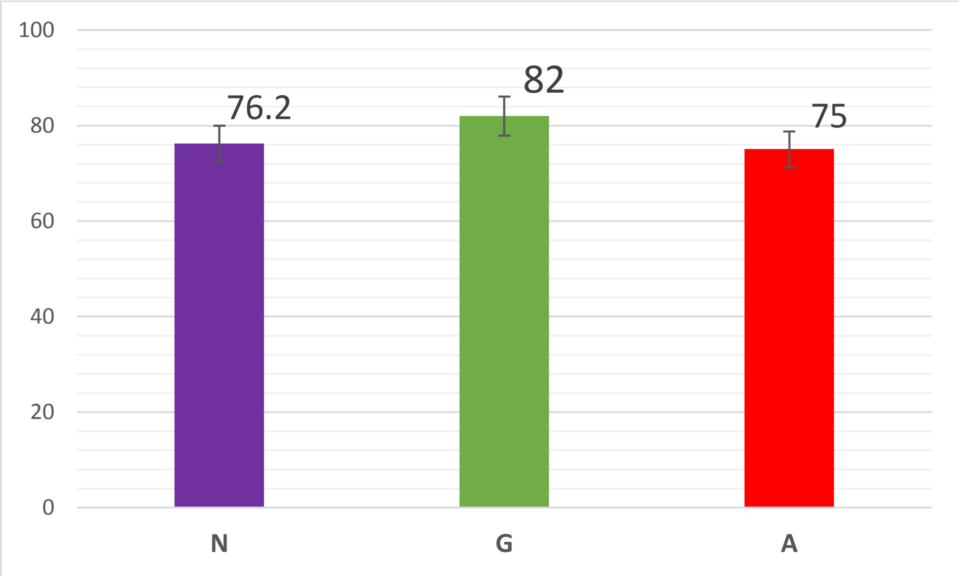

Figure 7.3 Section 2 results of the System Usability Scale

7.5. Engagement Questionnaire

The results of the Engagement questionnaire in section 1 trials showed all three conditions had slightly better than moderate level of engagement (see Figure 7.4). The average scores of each condition were 3.52 (*Md* = 3.68) for the None condition, 3.31 (*Md* = 3.2) for the Graphical Symbol condition, and 3.67 (*Md* = 3.58) for the UI Agent condition. Kruskal-Wallis test showed there was no significant difference between the conditions (H = 2.770, p = .250).

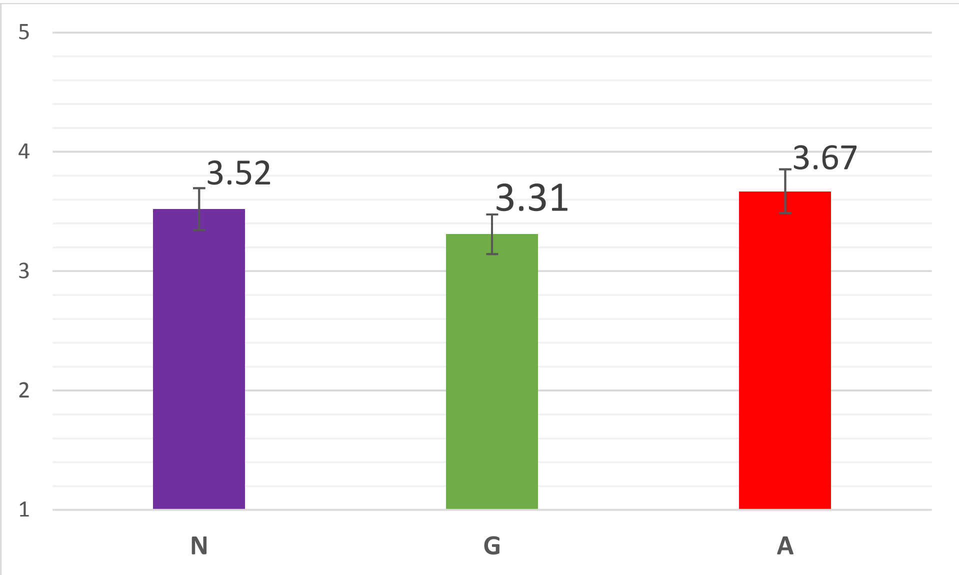

Figure 7.4 Section 1 results of the Engagement questionnaire

The results of the section 2 trials were similar to section 1 where again all three conditions show slightly better than average level of engagement (see Figure 7.5). The average engagement scores of each condition were 3.73 (Md = 3.84) for the None condition, 3.87 (Md = 3.87) for the Graphical Symbol condition, and 3.7 (Md = 3.84) for the UI Agent condition. Friedman test showed there is no significant difference between the conditions ($\chi^2(2) = 2.375$, $p = .305$).

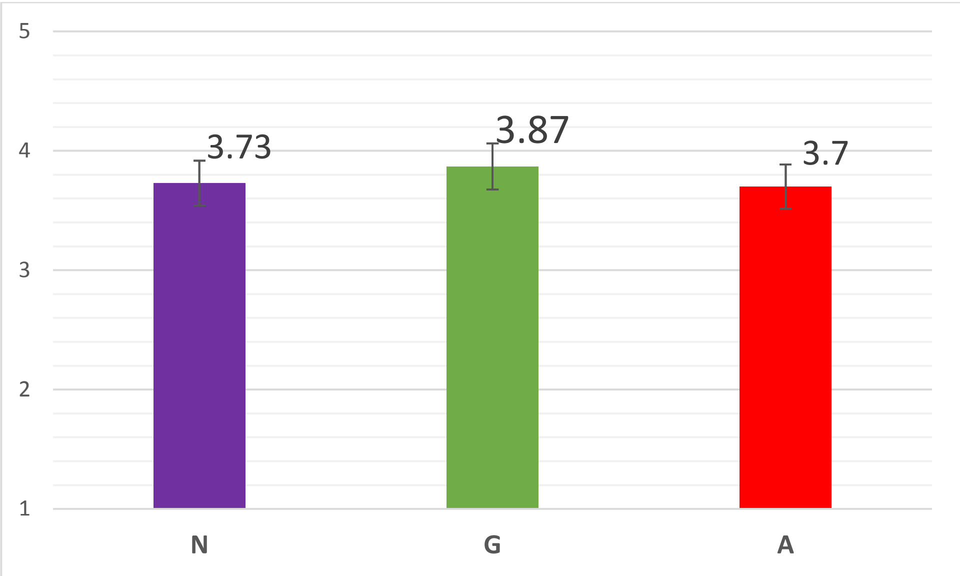

Figure 7.5 Section 2 results of the Engagement questionnaire

7.5.1. Ranking

In the post-experiment questionnaire the participants are asked to rank the three conditions based on their preference. Figure 7.6 summarizes the results where most of the participants ($N = 18$, 86%) ranked the Graphical Symbol condition in the first place, while more than half of the participants ($N = 12$, 57%) ranked the None condition at the last place. The ranking of the UI Agent condition was diverging between the participants with almost evenly split between the second and the third place. Friedman test revealed there is a significant difference between the conditions ($\chi^2(2) = 24$, $p < .0001$). Post hoc tests with Wilcoxon Signed Rank tests with Bonferroni correction ($\alpha = .0167$) revealed there was a significant difference between the Graphical Symbol condition and the others (vs. None: $W = -231$, $Z = -4.14$, $p < .0001$; vs. UI Agent: $W = -192$, $Z = -3.44$, $p = .001$), while no significant difference was found between the None and the UI Agent conditions ($W = -60$, $Z = -1.13$, $p = .259$).

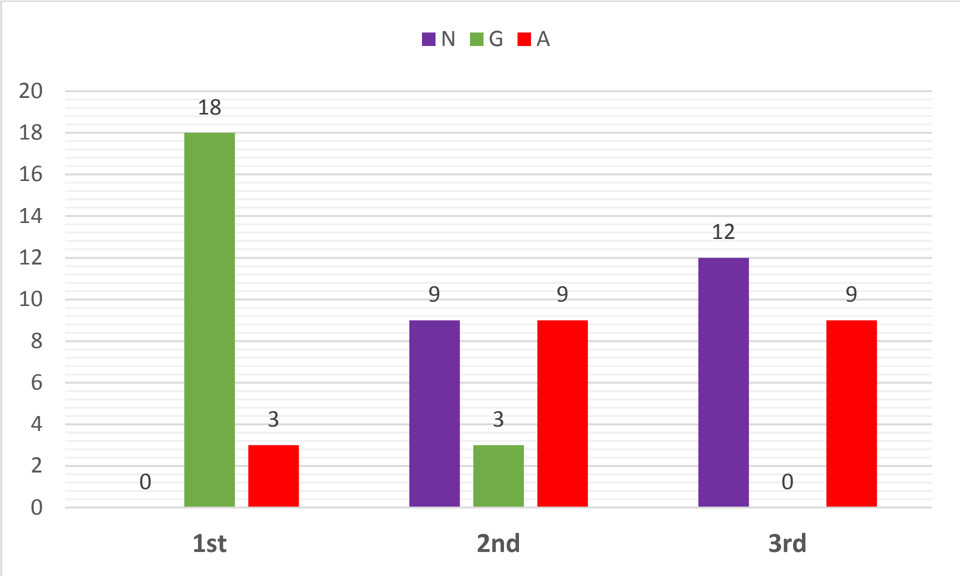

Figure 7.6 Results of ranking based on participants' preference

7.5.2. Observations, Interviews and Qualitative Feedback

During the experiment, the researcher carefully observed the participants and made note of their behaviors, such which hand used, if having any trouble at certain stage, or if showing smile on their face. The observations were compared among 18 participants, as for we did not had records of observation for the first three participants.

In section 1 of trials where we focused on the first encounter with the prototype system, we found those who tried the UI Agent condition smiling more when using the system for the first time (see Figure 7.7). While all the participants (6 out of 6) in the UI Agent condition showed smiling face, less than half of the participants smiled in the other conditions (3 out of 6 in the None condition, and 2 out of 6 in the Graphical Symbol condition).

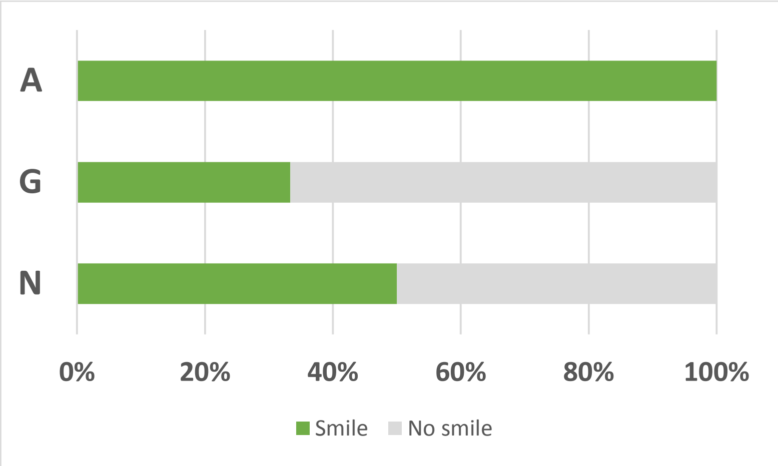

Figure 7.7 Smile on face on first encounter with the system

Other evidences of engagement with the UI agent character was also found in the second session, where few participants were found smiling when trying the UI Agent condition or even trying to play with it.

The prototype was considered as easy to use and figure out what to do from the first use. In the interview in the first section of the trials, the participants were asked if it was easy to guess what to do at the first place answering with rating on the scale of 0 to 10 (0: totally disagree ~ 10: totally agree). The median of ratings in each condition were 7 (IQR = [2-9]) for the None condition, 8 (IQR = [7-10]) for the Graphical Symbol condition, and 4 (IQR = [4-9]) for the UI agent condition.

When asked if it was easy to know where to stand, participants gave rating (*Md* [*IQR*]) of 9 [8-10], 9 [7-10], and 6 [4-8] to the None, Graphical Symbol, and UI Agent conditions, respectively. From observation we found about half of the participants in each condition did not understand the footprint symbol mainly due to incorrect depth cue with occlusion. A few of them (two each from the None and UI Agent conditions) tried to match their hands at the symbol as if they were needed to be selected as if with buttons, while most of them took some time to understand where they are supposed to stand at. In the UI Agent condition, two of the participants got distracted with the character trying to touch it and play with it.

When asked if it was easy to know how to select a button, those in the None condition gave relatively lower rating 7.5 [6-9], while the other conditions were rated 9[8-10] and 9[7-10] for the Graphical Symbol and UI Agent conditions, respectively. This is in line with the observations where two of the participants in the None condition having troubles such as trying to touch the screen or trying to use the head to select the button. Also, more participants in the None condition were observed trying gestures other than hovering and waiting to select a button. While about half (3 out of 6) of participants in both the Graphical Symbol and UI Agent conditions tried tapping or pointing for selection, all 6 out of 6 participants in the None condition tried various gestures other than waiting such as, tapping, double tapping, and pushing. In the interview, when asked to explain how to select a button, only half of the participants in the None condition described it in the correct way. Interestingly, two people in the UI Agent condition said that it is to hold the hand at the button until the bird arrives.

When asked about the relationship between the character and button selection in the final interview, about two third of participants (13 out of 18) described the relationship as the button gets selected when the UI agent arrives at the button, while two other participants did acknowledged the duck following user's hand but not sure about button getting selected, and yet other four replied there is no relationship between them.

Participants thought the UI agent character used in the prototype system was not much attractive. When asked if the character was attractive, participants rated with a median value of 7 (IQR = [3-8]) on the scale of 0 to 10. Two of the participants gave rating of 0 while none of them gave 10 out of 10. Most of the participants (13 out of 21) mentioned that they would not have rated the UI Agent condition higher even if the character was more attractive.

While all three conditions required the same amount of time (two seconds) to wait for selection after hovering over a button, somehow participants felt needing to wait longer in certain conditions. About two thirds of the participants (14 out of 21) stated there was a difference in time to wait for selection between conditions. Most (10 out of 14) of those who thought there was a difference felt the Graphical Symbol condition requiring the less amount of time. On the other hand, almost half of the participants (10 out of 21) felt the UI Agent condition took more time than the others.

7.6. Discussions

While the prototype system was accepted as relatively easy to use in the first encounter, the guide for guiding the user to stand at certain position needs improvement. About half of the participants had problem with incorrect depth perception due to incorrect occlusion between the user's body and the virtual footprint on the floor. This could be solved by implementing correct occlusion between real and virtual objects using depth masking technique.

The results showed subtle difference between the three conditions. Results of both System Usability Scale and Engagement questionnaires showed no statistically meaningful difference between the conditions. Yet participants ranked the Graphical Symbol condition significantly higher than others as the most preferred condition. Participants perceived the main strength of the Graphical Symbol condition as clear feedback on the time to wait when selecting buttons.

While not as clear as the timer animation in the Graphical Symbol, the UI Agent animation of bird pecking the button was still perceived as an indication of when the button gets selected. Noting that about two thirds of participants described the relationship between the character and button selection as in the manner it was designed for. While all of the participants had trouble figuring out how the button selection works in the first encounter, only half of the participants had similar problem with the UI Agent and Graphical Symbol conditions.

While two thirds of the participants understood the relationship between the UI agent and the button selection interaction, it was not clear to the rest of the participants. This could be improved by carefully deciding the visual designs of the buttons (or other UI elements) to reflect the relationship between the UI agent character and the buttons. For instance, a butterfly character could be matched with a button shaped like a flower.

The main benefit of using UI Agent appeared to be from the emotional perspective of the user experience, demonstrated by all of the participants in the UI Agent condition showing smiles on their face when trying the prototype system for the first time. It is also notable that how much the character is attractive does affect the overall user experience. While the character used in the study was not considered as much attractive, yet there were few participants trying to play with the character. While this could be a positive feature, as few participants mentioned, it could be also thought of as distracting in those applications that requires focus. This leads to a conclusion that the UI agents could be beneficial to certain types of application that requires user’s affection and engagement. On this note, entertainment applications would be one of the representative applications that could benefit from using UI agents.

8. Prototype System v3

Based on the results of the evaluation of the second prototype in Chapter 7, changes were made to the design. This chapter describes the changes made for building the prototype v3 including the design and implementation of the modified prototype system.

8.1. Theme

From the evaluation results of first prototype system, relationship between the UI agent and the theme of the application needed a strong relation. For this, we chose the theme as fruit quiz where a fruit or a vegetable needs to be fed to the UI agent which is a parrot character. Parrot character is chosen because it is known to like fruits, vegetables and nuts. The quiz questions are given by showing a clip art image of a slice of fruit or a vegetable and the user is required to select an answer whether it is fruit or vegetable.

The main role of the UI Agent (i.e. the parrot character) in the prototype system is to guide the user for interacting with the system. This includes providing hints to the user to stand at the right place in the engagement stage, and giving hints on how to select an answer using gestures.

The summary of 3 changes that were made.

1. For the initial user engagement we included occlusion.
2. For the main application, instead of buttons, UI agent is used.
3. Instead of duck character, a parrot character is used to have a strong relationship with the type of application and attractiveness. Figure 8.1 shows an image of the virtual character used, a Parrot. Based on the virtual character to be used in the prototype system, we chose the theme of the quiz around the topic of fruits and vegetables.

Figure 8.1 A parrot virtual character used for the final prototype system

8.2. Initial Engagement Phase

Figure 8.2 Prototype system at the idle stage

To reflect this new theme, in the idle stage, splash screen and virtual character are changed. As the screen unfolds, the UI agent starts providing visual cues to the user as shown in the Figure 8.1.

8.3. Depth Occlusion

As per the evaluation results of previous prototype, depth cues are provided through occlusion in order to identify the cues better in real world space. In place of footprint symbol, a simple rectangular plane is shown.

The functionality retained same as in the previous prototype. Once a user expresses his or her interest through standing in front of the prototype system and facing towards the screen, the system should guide the user to stand at the ideal place for further interaction with it. Occlusion is applied to the user image stream by drawing invisible cylinders on his body. Whenever user covers the visual cues (see Figure 8.3), they are drawn with user image stream to provide an illusion that the cues are behind or in front of the user (see Figure 8.3).

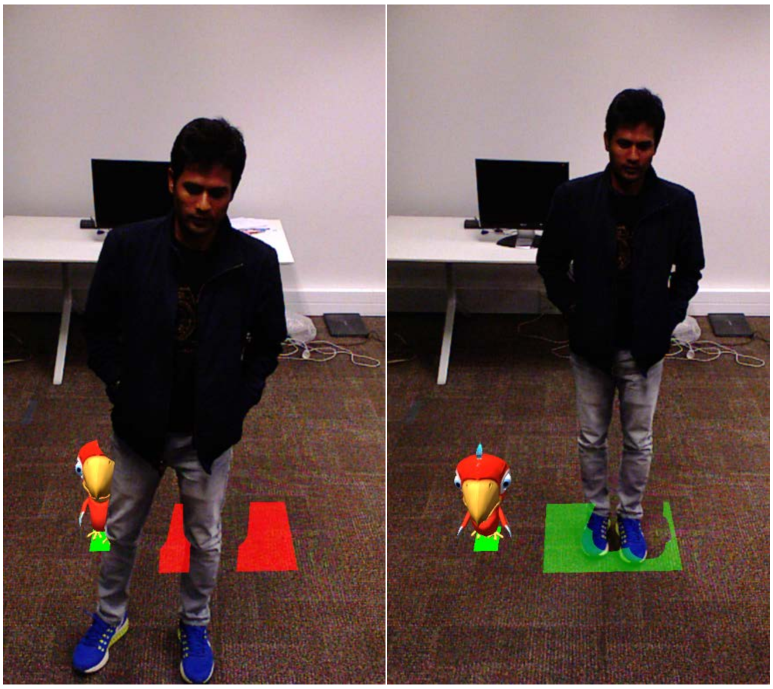

Figure 8.3 Guiding a user to stand at the ideal place in the engagement stage

8.4. UI Agents as Buttons

Based on qualitative feedback from previous prototype, stronger relationship was required between the kind of application and the UI character. To meet this requirement, the main application is designed to be a fruit or vegetable quiz with UI agent as a parrot.

Compared to the previous prototype design, in the third prototype system we used the UI agents as buttons instead of cursors. From the qualitative feedback (See section 7.5.2), Majority participants button selection in UI agent condition is slower and button selection works when UI agent reaches the button. To overcome this, UI agents are used as buttons instead of following user's hand so that users get faster feedback and intuitive selection process using UI agent animations. In place of buttons, a user agent is used with relevant static graphics (see right side picture in Figure 8.4. Another change that was made is to show the slice of a fruit or a vegetable as a cursor on top of leading hand so that users see a strong relationship for the gesture. For every question, a smaller image of fruit or vegetable that is same as in the question will be displayed on top the leading hand (see right side picture in Figure 8.4). This provides an analogy of feeding a bird with the fruit or vegetable by holding the user's hand near the bird.

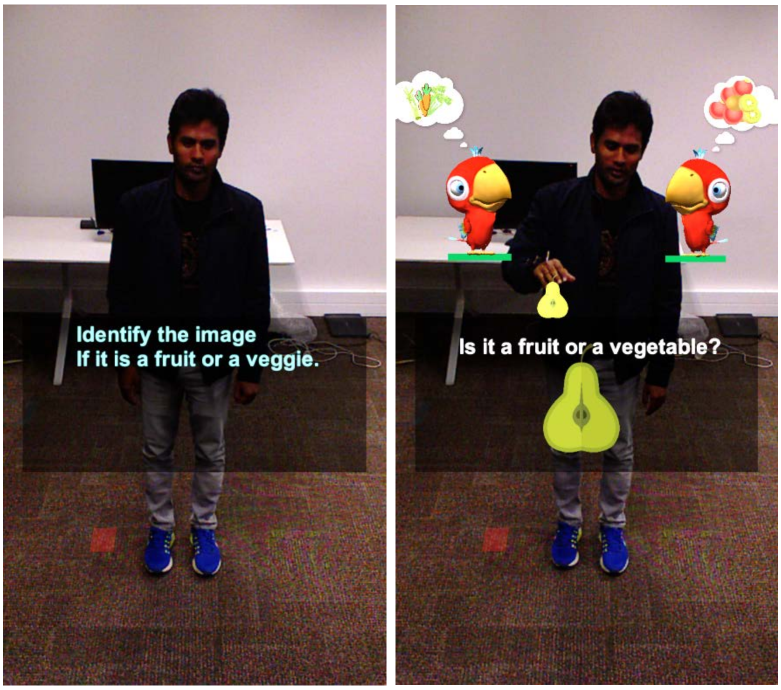

Figure 8.4 Instruction and quiz screens

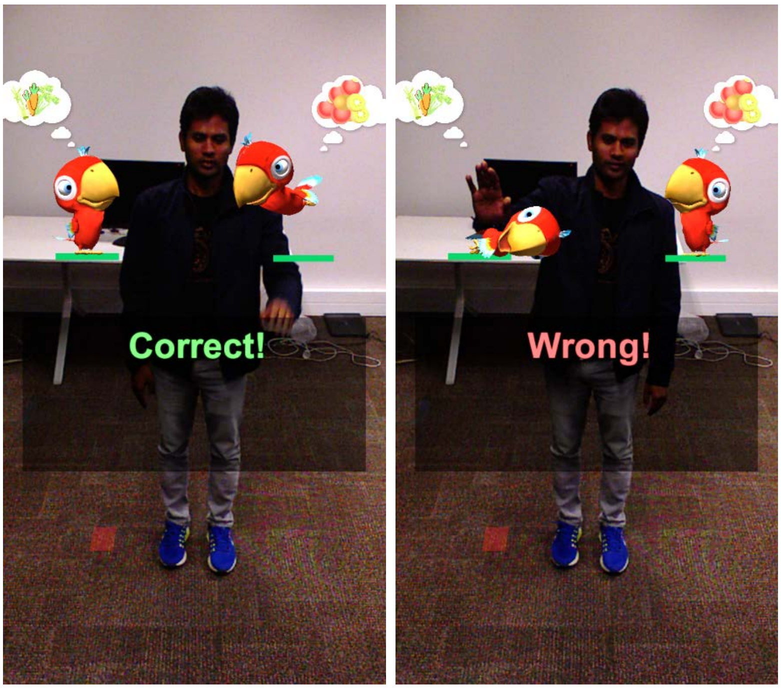

Figure 8.5 The screen showing whether the answer is correct or wrong

UI agent provides the feedback and helps the user with the kind of gesture to be used. As the GUI button provides a timer feedback to let users to hold their hand on the button for hovering gesture (see Figure 8.6), the UI agent as a button plays an animation of eating the fruit or vegetable to indicate that the UI agent has recognized the user's hand on the it and to suggest holding the hand for a while (see Figure 8.7).

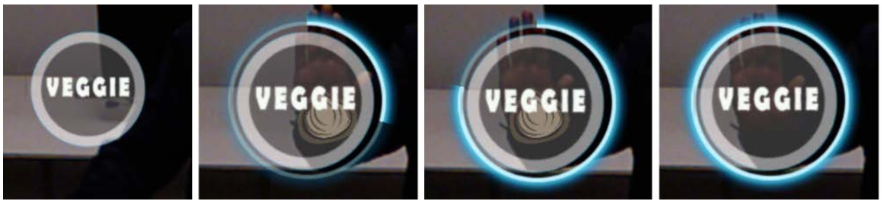

Figure 8.6 Button animation for Graphical condition

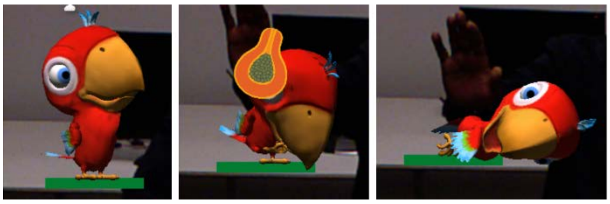

Figure 8.7 Button animation for selection in UI Agent condition

Once the user selects an answer, the answer buttons disappear and the screen shows whether the chosen answer is correct or not (see Figure 8.5). After a few seconds, the prototype system shows the next question for the quiz and repeats until all of the quiz questions are answered. After answering all of the quiz questions, the system finally shows a 'Thank you!' message, and the system turns back to the idle stage after few seconds.

9. Evaluation of prototype v3

To evaluate the proposed user interface design, we conducted an experimental user study with the revised prototype system. The goal of the user study was to evaluate the effects of different visual cues on user's experience while interacting with the virtual mirror interface using gestures. The user study design was reviewed and approved by the Human Ethics Committee at the host institution according to Low Risk process. This chapter describes the details of the experimental design, report on the results, and discuss on the implications of the findings from the user study.

9.1. Setup

Same setup is used as in previous prototype evaluation (See 7.1 Setup).

9.2. Experimental Design

The user study was designed as a formal experiment, with the main independent variable as the design of visual cues used with the prototype system. We compared two types of visual cues: Graphical Symbol, and UI Agent. The Graphical Symbol condition (G) the prototype system showed visual cues of graphical symbols on guiding the user what to do or how to interact with the application (See Figure 8.6). In the UI Agent condition (A), the proposed method of using UI agents as visual cues was applied to guide the user with interaction (See Figure 8.5). Table 9.1 summarizes the difference between the conditions in more details. (For more detailed design of the visual cues, please read chapter 8)

The task was to use and interact with the prototype system based on what is shown on the screen. While we used within-subject design for investigating the Main Application stage (answering quiz questions on fruits and vegetables), we chose between-subject design for the Engagement stage as the main focus of the investigation is around intuitiveness which requires participants to be new to the system (i.e. never used it before). The engagement stage compared the two conditions of with or without correct depth occlusion

Table 9.1 Visual cues in the experimental conditions

Stage	Graphical Symbol	UI Agent
Idle	- Full screen image background - A hole of video stream highlighting user's face	- Full screen image background - A hole of video stream highlighting user's face
Engagement	- Red rectangular plane highlighting a zone in real world space - UI agent showing hints on where to stand	- Red rectangular plane highlighting a zone in real world space - UI agent showing hints on where to stand - Occlusion for bird and rectangular plane where user is expected to stand.
Main Application	- Button highlighted on hover - Button glows brighter on selection - Timer indicator	- Fruit or vegetable image shown in question follows the hand as cursor - UI agent tries to eat the fruit or vegetable that appears on top of the hand when his/her hand reaches the UI agent location

9.2.1. Hypothesis

In the first section of the experiment we have one hypothesis

H_1 – There is a significant difference in task completion time between UI agent with occlusion and UI agent without occlusion conditions.

In the second section of the experiment, we have three hypothesis.

H_2 – There is a significant difference in system usability scale between the graphical symbol condition and UI agent conditions.

H_3 – There is a significant difference in user engagement between the graphical symbol condition and UI agent conditions

H_4 – There is a significant difference in number of wrong gestures between the graphical condition and UI agent conditions.

9.2.2. Procedure

The user experiment followed the procedure summarized in Table 9.2. We followed the script (see Appendix L) for dealing with the participants in consistent manner. An experimental session started with welcoming a participant to the experimental environment. The participant was then asked to sit and read the information sheet (see Appendix C), and sign the consent form (see Appendix D). After signing the consent form, the participant was asked to fill in the pre-experiment questionnaire (see Appendix E) which asked demographic information and background of the participant, before continuing on to the experimental trials.

The first section of the experimental trials was for investigating on how intuitive is the system for the first time users. Participants were asked to stand in front of the prototype system and interact with it only based on what is shown on the screen but without any further instruction from the researcher. In this trial, the system started from the Idle stage and continued on up to selecting the Start button in the Main Application stage. As this section was in between-subject design, each participant experienced only one of the experimental conditions. After finishing the task, participants were asked to sit at the table and answer the per-trial questionnaire (see Appendix H).

Table 9.2 User Evaluation Study Procedure

Duration	Procedure
5 min.	- Welcome - Informed consent - Pre-experiment questionnaire
Section 1 (between-subject design)	
5 min.	- Initial trial of the prototype system in one condition - Per-trial questionnaire - Interview
Section 2 (within-subject design)	
10 min.	Repeat for two condition: - Trial of the prototype system in a condition - Per-trial questionnaire
5 min.	Post-experiment questionnaire
5 min.	Interview & debriefing

The second section of the experimental trials was in within-subject design. The participants had two trials of using the prototype system in different conditions. The order of conditions was counter balanced. In each trial, the system started from the Engagement stage where it indicates the participant where to stand. After standing at the position, the participant continued on to the other stages following the information provided on the screen. Six quiz questions on the topic of fruits and vegetables were asked in each trial where the participants had to answer by selecting the UI agents in UI agent's condition and "Fruit" and "Vegetable" buttons in Graphical symbol conditions. (See chapter 8 for more details). Different set of quiz questions were used in each trial. After answering ten questions the system showed a ‘Thank you!’ message for concluding the trial. After each trial, participants were asked to answer the per-trial questionnaire (see Appendix H). After finishing both the conditions, participants were asked to answer the post-experiment questionnaire (see Appendix G). Finally, the participants were interviewed briefly before concluding the experiment.

9.2.3. Measurements

The main method of measurement we used was using questionnaires. For the initial user engagement section we used task completion time through logs and qualitative feedback about the experience. This part is also video recorded to get qualitative data and also as fallback method to measure task completion time. For second section which is the main application, for each trial we used System Usability Scale (SUS) [40] [41] and O’Brien’s Engagement questionnaires [42] to measure the usability of the prototype system and the user’s level of engagement. The SUS questionnaire results were aggregated into the range of 0~100, while the Engagement questionnaire results used the range of 1~5. A semi-structured interview was conducted in each section of trials, asking various questions such as how easy it was to know how to interact with the system, any perceived difference in the time delay for selection, and how did the UI agent character affected the user experience.

In addition to subjective measures on usability and user experience, we also collected objective measures through observation, such as number of time the participant tried wrong gestures for selecting a button and the type of gestures tried, which hand was used for interacting with the system, and if the participant had troubles in using the system.

9.3. Participants

We recruited 24 participants through advertising on the university campus and online community websites. None of the participants from previous experiment were chosen to participate in this experiment to avoid previous knowledge. The participants were between 18 to 63 years old (*Mean* = 24.3, *Standard Deviation* = 9.49, *Median* = 21) and 9 of them were female (37.5%). Most of the participants used right hand as their dominant hand (*N* = 18, 75%) while four participants answered that they use both hands, and two participants answered the left hand.

In the pre-experiment questionnaire (see Appendix E), when asked how frequently they use public information systems, more than half of the participants (N = 13, 53.2%) answered a few times a month, but there was none who answered ‘every day’.

The participants were not much familiar with gesture based interfaces, as when asked if they have used gesture interface before, most of the participants (21, 87.5%) answered less than few times a year or not at all. When asked if they have played Microsoft XBOX Kinect motion games before, 9 of the participants (37.5%) answered they have not played at all, while more than half of the participants (N = 12, 50%) answered few times a year. The response was similar when asked if they have played Nintendo Wii or Sony MOVE, where 12 of the participants (50%) answered not at all, and 10 of the participants (41.7%) answered few times a year.

When asked if they have used Augmented Reality (AR) app or interface before, 8 of them (33.3%) answered they were not aware of what AR is, while the rest (N = 16, 66.6%) had used at least few times a year (the mode was ‘I am not aware what AR is’ as 8 participants has chosen this answer).

There were three 7-point Likert item questions (1: Strongly disagree ~ 7: Strong agree; 4: neutral) in the pre-experiment questionnaire, which showed participants saw themselves moderately using gestures in everyday life (*Median* = 5, *Inter-Quartile Range* = [3.25-6]), positive about characters (*Median* = 5, *IQR* = [4-6]), and they reported that their favorite characters moderately affects their experience with a product (*Median* = 5, *IQR* = [3.25-6]).

9.4. Results

Here we report the results of the task completion time for the first section of the trials, and next the results of the per-trial and post-experiment questionnaires of the second sections of the trials, then summarize the findings from the interviews and observations. The per-trial questionnaires (System Usability Scale and Engagement questionnaires) results are reported for each sections of the experimental trials. The first section (section 1) of trials focused on the first time use of the system. This was a between-subject design. While the second section (section 2) was within-subject design focusing on the main application. See section 9.2 for more details of the experimental design. All inferential statistics were tested with an alpha level of 0.05, unless it is stated otherwise.

9.4.1. Task Completion Time

For section 1 trials (N=24), the condition that we tested was whether occlusion for UI agent and Foot-mat marker help the participants to stand at the designated space in the real world.

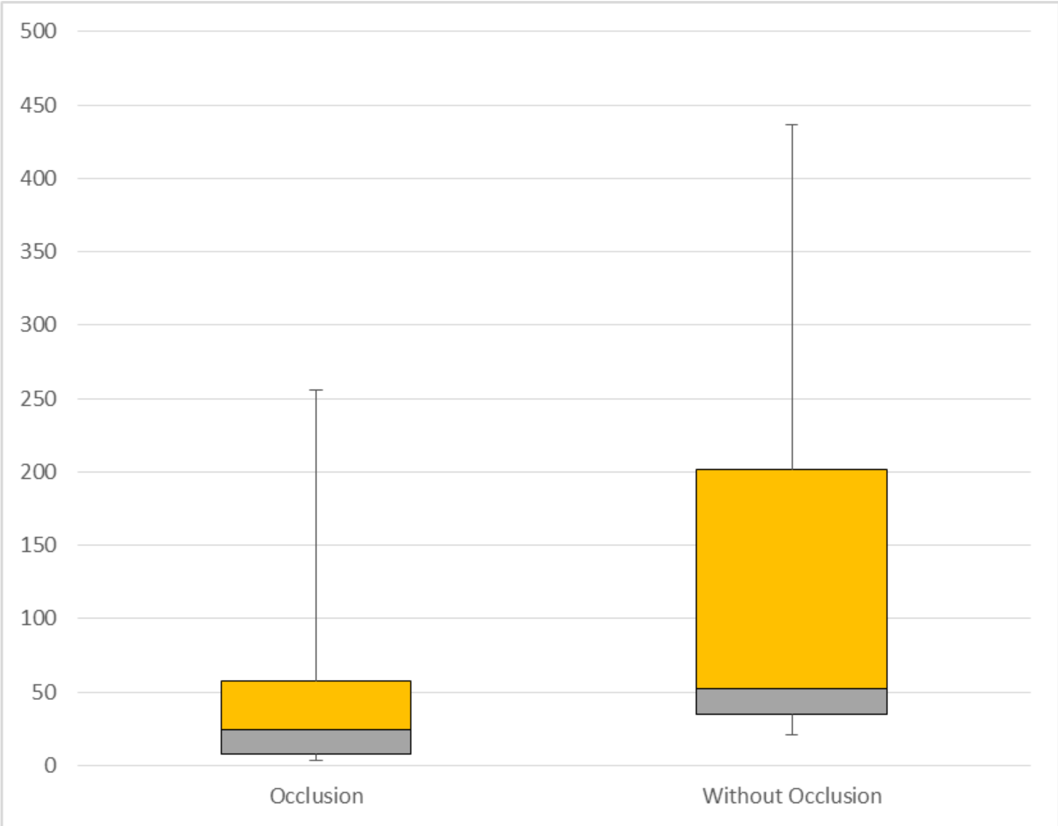

Figure 9.1 Box plot of task completion time

One sample in the occlusion condition was marked as an extreme outlier (261 Secs) and it was removed from the data. After the removal of the outlier, the average task completion times were 30.54 sec (Md = 27 sec) for "with occlusion" and 152.5 sec for "without occlusion" (Md = 88.5 sec) . Data was further analyzed for normality using Shapiro-Wilk test (Condition 1: $W = .644$, $p < .001$ and Condition 2: $W = .828$ $p = .020$) and found that it was not normally distributed. Hence, Mann-Whitney U test was conducted on the data and found statistically significant difference ($Z = -2.369$ $p = .018$) between task completion time with occlusion compared to without occlusion.

9.4.2. SUS Scores

The results from the section 2 trials show both of the conditions having slightly above the average usability level (68). The mean SUS scores of each condition were 78.12 (*Md* = 81.25) for the Graphical Symbol condition, and 74.27 (*Md* = 75) for the UI agent condition (see box plot in Figure 9.2). There was no significant difference found between the conditions based on Wilcoxon Signed Ranks test (Z =.4, p = .689).

When conducted Wilcoxon Signed Ranks test between the graphical symbol condition SUS score and an average SUS score of 68, there was significant difference found (Z =.-2.745, p = .006) which shows that graphical condition score is above average from Figure 9.3.

Wilcoxon Signed Ranks test between the UI Agent condition and an average SUS score of 68, there was no significant difference found (Z =.-1.801, p = .072) which shows that it is around the average score of 68 from the Figure 9.3.

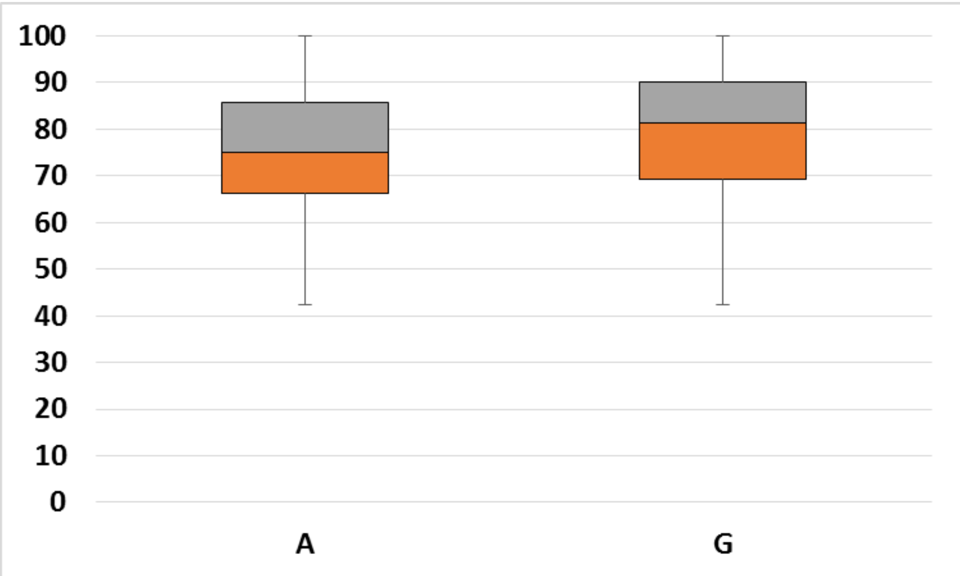

Figure 9.2 Box plot for SUS scores

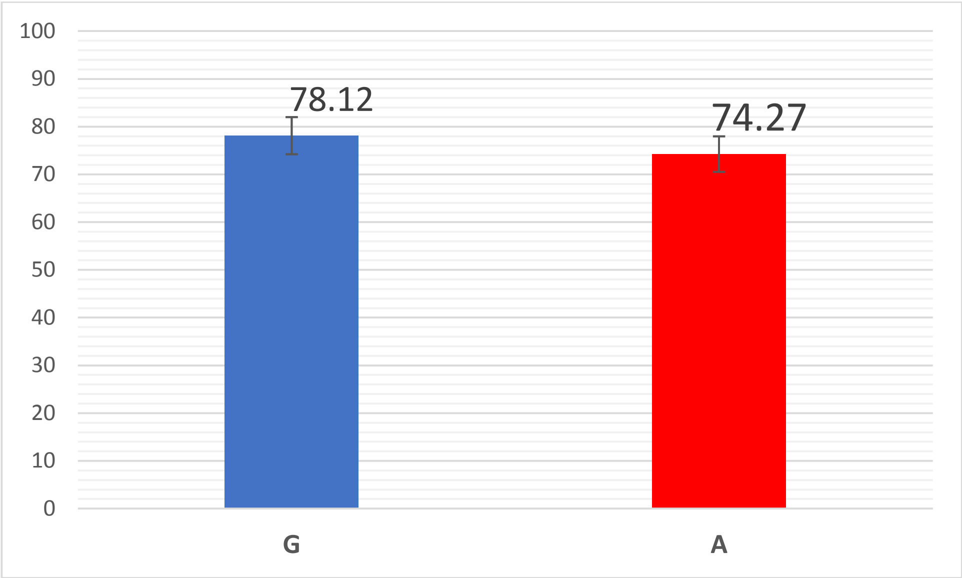

Figure 9.3 Section 2 results of the System Usability Scale

9.4.3. Engagement Questionnaire

The results of the trials in section 2 show that both of the conditions show slightly better than average level of engagement (see Figure 9.4). The average engagement scores of each condition were 3.69 (Md = 3.65) for the Graphical Symbol condition, and 3.9 (Md = 3.85) for the UI Agent condition. Wilcoxon Signed Ranks Test showed there is significant difference between the conditions ($Z = -2.334$, $p = .020$).

When user engagement levels are compared with a middle value of 3 for each condition, Wilcoxon Signed Ranks Test showed that there is statistically significant difference between middle value and Graphical symbol condition ($Z = -4.286$, $p < .001$) and also for UI Agent condition ($Z = -4.058$, $p < .001$).

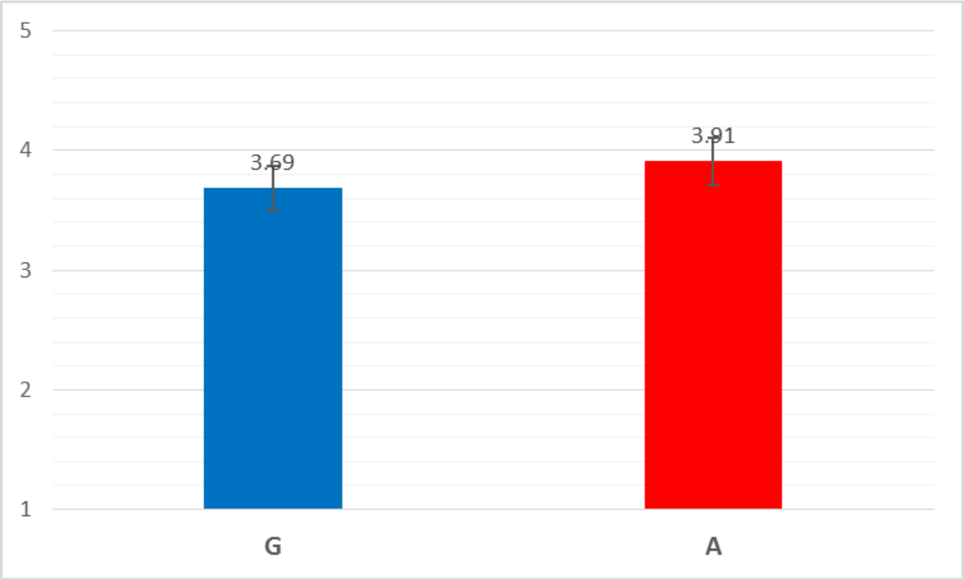

Figure 9.4 Section 2 results of the Engagement questionnaire

Going further into different factors of user engagement shows the following:

The average scores for Focus subscale of user engagement for each condition were 3.13 (Md = 3.0) for the Graphical Symbol condition, and 3.4 (Md = 3. 5) for the UI Agent condition. Wilcoxon Signed Ranks Test showed there is significant difference between the conditions (*Z* = -2.334, *p* = .020). Similarly, significant differences were found (see Figure 9.5) for Novelty (Z=-2.72, *p*=.006), Endurability (Z=-1.89, p=.058) and Aesthetics (Z=-2.15, *p*=.032).

In contrary, no significant difference was found among Involvement (Z=-1.040, *p*=.020) and Usability subscales of user engagement (Z=-.061, *p*=.951).

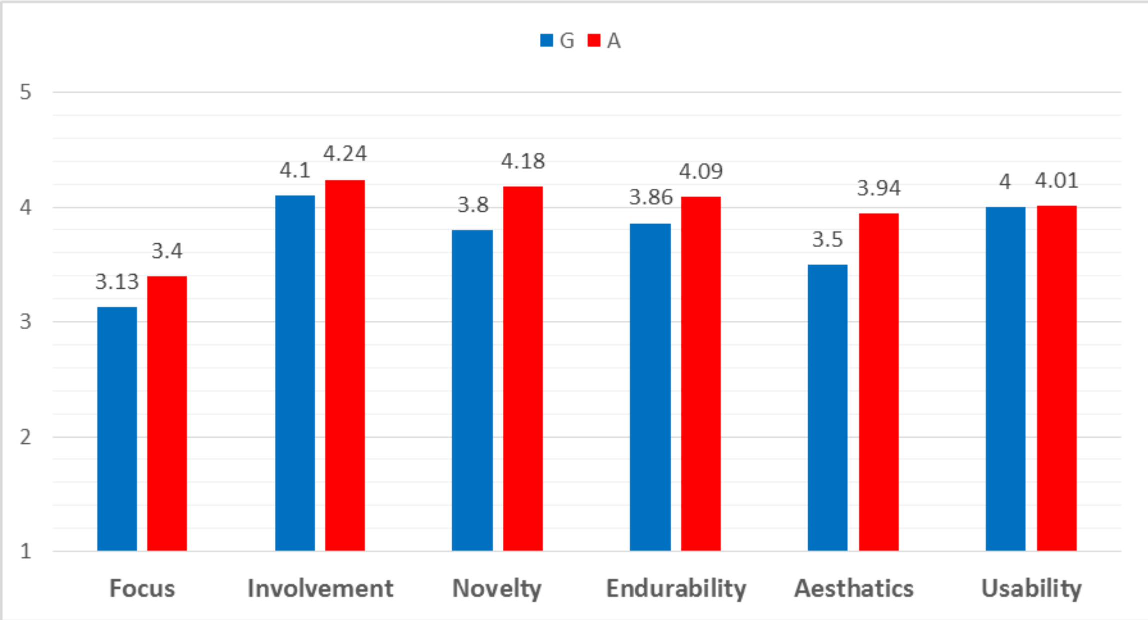

Figure 9.5 Section 2 results of the Engagement questionnaire - subscales

9.4.4. Ranking

In the post-experiment questionnaire the participants are asked to rank the two conditions based on their preference. Figure 9.6 summarizes the results where more participants ranked the Agent condition (N = 14, 58.3%) in the first place as compared to the Graphical condition (N = 10, 41.7%). Wilcoxon Signed Rank Test revealed that there is no significant difference between the conditions (Z = -.816, p = .441).

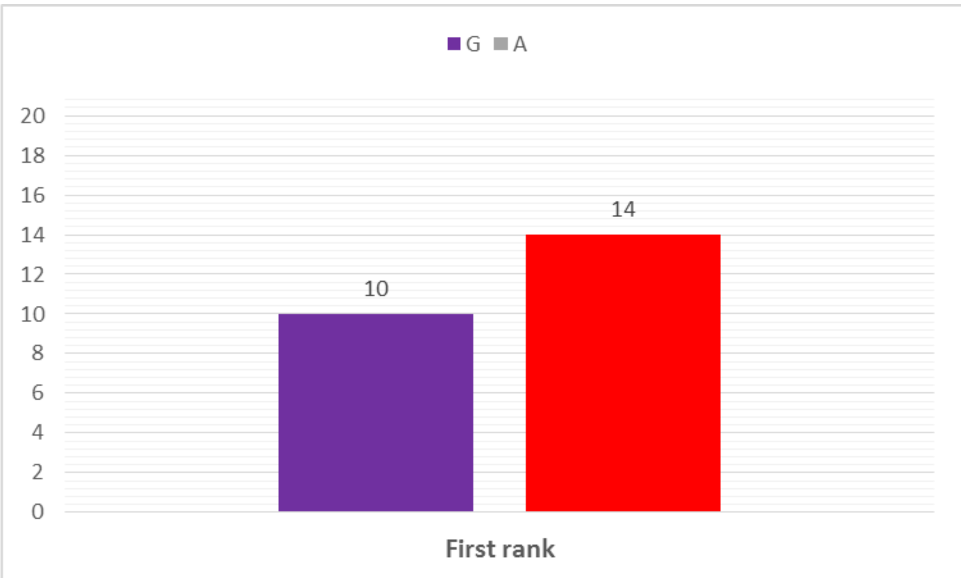

Figure 9.6 Results of ranking based on participants' preference

9.4.5. Errors

The results of section 2 trials show that there were considerably less errors made while using UI agent condition. Figure 9.7 shows the average number of errors made in each condition.

The average errors made during gestures were 0.67 (Md = 0) for graphical condition and were 0.25 (Md = 0) for UI agent condition. Data was further analyzed to check for normality using Shapiro-Wilk test and found that both the conditions data were not normally distributed (Condition G: $W = .717$, $p < .0001$ and Condition A: $W = .531$ $p < .0001$). Hence Wilcoxon Signed Ranks Test was conducted and found significant difference ($Z = -2.178$, $p = .029$) As the mean value of the GUI condition is more than twice of that in UI Agent conditions (See Figure 9.7), we can conclude that there was a significant reduction in errors made in UI agent condition.

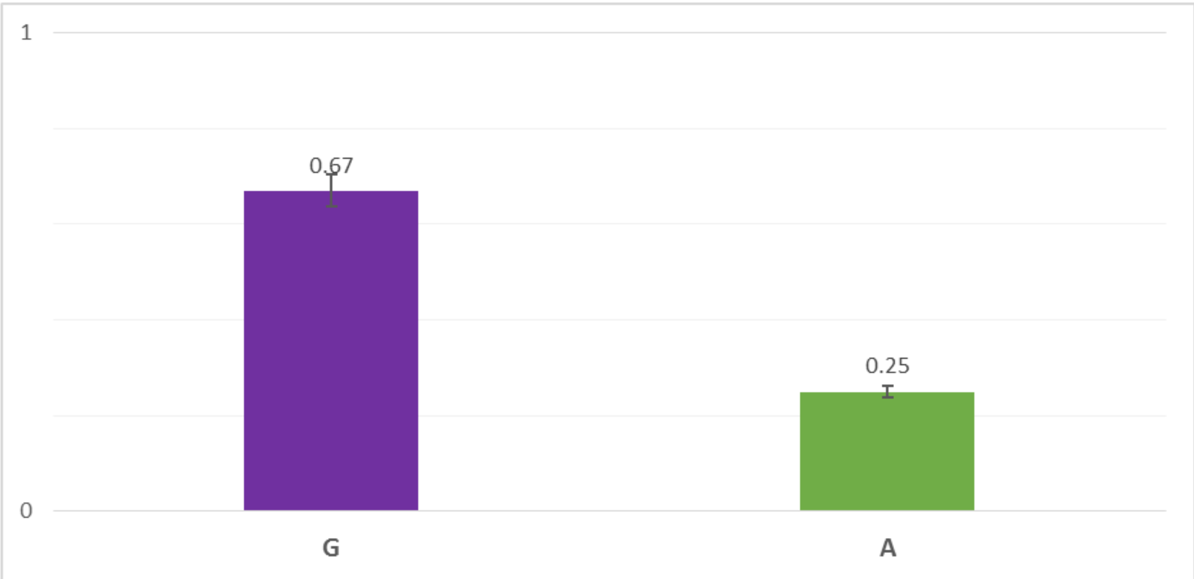

Figure 9.7 Results of number of errors committed

9.4.6. Observations, Interviews and Qualitative Feedback

During the experiment, the researcher carefully observed the participants and made notes of their behaviors, such as which hand was used, if they were having any trouble at certain stages, or if they were showing smile on their face. The observations were compared among 24 participants.

Some of the qualitative data that was collected for the section 1 of the experiment, 3 participants commented saying that they would like to see foot prints on red color rectangular plane which was used as a visual cue to guide users where to stand. Interestingly, we removed the footprints from experiment 1 and replaced it with a rectangular plane. As seen in the previous experiment, most people tried to play with the UI agent showing positive engagement with it.

While both the conditions required the same amount of time (two seconds) to wait for selection after hovering over a button, many participants (17 out of 24) had different perceptions about waiting time in different conditions. Most of the participants (10 out of 24) answered that the UI Agent condition required less waiting time to select an answer, whereas the other 7 participants believed that Graphical condition required less waiting time to select the answer. Interestingly 7 participants out of 24 responded that they did not notice any waiting time difference for selecting an answer between both the conditions.

Participants rated UI agent character used in the prototype system as attractive. When asked if the character was attractive, participants rated with a median value of 7.1 (IQR = [7-8.75]) on the scale of 0 to 10, with the lowest ranking being 1 and highest being 9. Most of the participants (17 out of 24) mentioned that they would not have rated the UI Agent condition higher even if the character was more attractive.

Wrong gestures appeared in the GUI condition which included tapping the button (5), insufficient holding times (4), waving or swiping (3), pointing their fingers (2), and grabbing the button (1). On the other hand, errors committed in UI Agent condition included hovering their hand (3) over the UI agent for an insufficient duration (2 seconds) and hovering their hand over static image of fruits and vegetables (1) instead of UI agents.

When asked to pick an interface that was fun, an overwhelming number of participants (18 out of 21) said that they had more fun while using the interface with UI agents. Only two participants said that they had fun with graphical symbol condition and one participant said that he had fun with both the interfaces.

9.5. Discussions

Overall the prototype system was accepted as very easy to use in the first encounter. Participants were able to follow the occlusion depth cues with ease. Task completion times for depth cues were almost 5 times longer than that of with occlusion.

System usability of both the interfaces showed no statistically significant difference. The Graphical Symbol condition was rated above average while the UI Agent condition was rated around average. This is possibly due to the more visual information in the UI Agent condition. Participants also felt that the interface could have been simpler by reducing the visual elements. Two of the participants who rated the UI Agent interface SUS scores as the lowest had a problem reaching the buttons (UI Agents). It is worth noting that from informal observations that they were attempting to feed the food to the beak of the parrot, which is a part of the bird that is higher than the original button. They could be clearly observed to be attempting to reach out to a position that was uncomfortable for them. This coupled with the insufficient feedback from the system to move and readjust the buttons according to participants' heights could have resulted in their dissatisfaction in the interface.

User engagement was higher in the UI agent condition with significant difference compared to Graphical Symbol condition. There was a significantly higher user engagement for the subscales of Focus, Novelty, Endurability, and Aesthetics for the UI Agent condition as compared to the Graphical Symbol condition. There is no significant difference in usability subscale, which is in line with the SUS scores. Although there was a higher level of involvement in the UI Agent condition, there was no significant difference found.

Participants' preference for UI agents interface was also reflected in terms of ranking for fun. More than 90% of the participants ranked the UI agent interface as the most fun interface and majority of the participants rated it as their most preferred interface. Two thirds of the participants stated that graphical condition was easy to understand. Some participants preferred the Graphical Symbol condition due to its simplicity, and some participants who preferred the UI Agent condition liked it for its feedback and fun factor.

Interestingly, although both conditions required the same amount of selection waiting time of two seconds, participants had varying perceptions of the amount of time they felt they have waited. Most of the participants believed that the UI Agent condition required less waiting time than the Graphical Symbol condition. This could be due to the participants having more fun in the UI Agent condition, which led them to believe less time had passed. Another reason for this could be that there were more visual elements to focus on which removed boredom whilst holding up their hands.

Also, gestures were performed more accurately in UI agent condition and a significant difference was observed when compared to the Graphical Symbol condition. More incorrect gestures were made in the Graphical Symbol condition than in the UI Agent condition. Most of the incorrect gestures made in the Graphical Symbol condition were due to unclear visual affordances while a majority of the incorrect gestures made in the UI Agent condition were due to insufficient feedback.

In the interview, it was found that most participants did not believe that the attractiveness of the character had an impact on their rating. This result was found to be in line with the results from the pilot study. However this result was subjective and could be an area of exploration for further studies.

The use of UI Agents in an interface was found to improve user engagement in terms of focus, novelty, endurability, and aesthetics, as well as increase the fun factor and likeability for users. UI Agents also showed an improvement in the accuracy of performed gestures, which could demonstrate the improved interpretation of the visual affordances. This could encourage users to interact more with public information display interfaces. Although user engagement was improved, the overall usability appeared to have dipped when compared to a simple Graphical Symbols interface due to the increase in visual elements on the screen. Due to the higher level of user engagement and fun factor, UI Agents are more suitable to be used in education and entertainment related interfaces.

10. Conclusions

This research investigated the feasibility of using UI agents as visual affordance cues in gestural interfaces with large-screen displays. The research examined the application of the gestural interface to public information displays, which demand more user engagement. I developed a prototype system to conduct a series of user studies that evaluated various aspects of the participants' experience. From the qualitative and quantitative results drawn from this research, the key points are as follows:

- More people smiled with the UI agent condition during initial engagement phase of interaction.

- In the initial phase of testing whether the prototype engages the user, the use of UI agents resulted in higher levels of engagement when visual cues were presented in the real-world space

- Correct depth occlusion cues helped users to understand how to interact with the visual cues more easily.

- The reception of UI agents depended largely upon the relationship between the purpose of the application and UI agent that is used. For instance, users understood intuitively that they should feed the fruit or vegetable to the parrot (UI agent), whereas when there is no relationship between the bird (UI agent) and the button in prototype v2, users did not understand the need for the bird to be in the interface and therefore did not like it.

- The presence or absence of the UI agent did not significantly affect the usability of the interface.

- The UI agent, however, did encourage users to engage with the interface, particularly in terms of novelty, endurability, and aesthetics (metrics as defined under HCI design, reference). An overwhelming number of participants found interaction with the UI agents interface to be more fun than without the UI agents.

- Users were faster to learn the gestures and more accurate when interacting with UI agents.

- When asked to rank the interfaces, more users preferred interacting with the UI agents than with the graphical symbol interface, although this finding is not statistically significant.

10.1. Limitations

This section describes some of the key challenges faced in this study. These could be addressed in future studies.

- During the first section of the interface (testing initial user engagement), participants reported that they could not understand what the UI agent was meant to convey. This report is likely because the range of animations was limited. In addition, participants wanted to play with the UI agent, and they expected the UI agent to respond to these actions. Therefore, future research should use more precise visual cues/animations in order to show the limitations on the UI agent's ability to interact with the user.

- In this study, the interface did not respond to the height of the participant. In particular, the height of buttons was out of reach for a few participants who could not reach the beak area of the parrot (UI agent). Participants correctly assumed that they were to feed the fruit or vegetable to the parrot in order to select their answer. Therefore, buttons and/or UI agents should be designed to automatically adjust based on the participant's height.

- All visual cues are designed to be subtle while expecting users to figure out the action required. Based on feedback from users, UI agents in the main application should be able to guide users when there is no activity from the user, in order to increase the engagement.

11. Future Work

In this research, we studied the use of visual affordances using UI agents for gesture interaction with large-screen displays. This is the first step towards usage of UI agents to build more engaging gestural interfaces. From the conclusions, the following are some areas that need further research:

- This study explores the use of the “hover to select” gesture. As more participants learnt the gesture without any training and with minimum number of errors, other gestures such as tap, double tap, drag and drop, shake or new type of gestures could be further explored.

- Further research is required on whether reducing the visual elements increases the overall system usability.

- In this study we explored using UI agents as proactive cursor in the initial high-fidelity prototype and in place of buttons in final prototype. In both the prototypes, we focused on two choices for the user to select. Further research is required to test whether it is viable to use UI agents as buttons when more options have to be provided to the user. For example, a virtual shopping application would require more than two actions or choices to choose from. In this case, effectively using UI agents without filling up more screen space is a challenge for future researchers.

- Pairing audio cues with UI agents would be another potential area of exploration for future research.

- In our study, we applied interactive objects in the real-world space using depth cues and UI agents. We used a large-screen 2D display, and future research could use the same system using a 3D stereoscopic displays and different visual cues.

- Future research could apply the findings from this research to improve the proposed prototype system. Additionally, further research could investigate the use of UI agents to guide interactions besides gesture-based button selection.

- Future research can also check the type of user agents and its impact on different gestures so as to find out the types of user agents that suit different gesture types. Primarily, this study focuses only on implementation in 2D interfaces. Further research can investigate on the use of UI agents in 3D interfaces.

References

[1] M. Billinghurst, A. Clark, and G. Lee, “A Survey of Augmented Reality,” *Found Trends Hum-Comput Interact*, vol. 8, no. 2–3, pp. 73–272, Mar. 2015.

[2] S. White, L. Lister, and S. Feiner, “Visual hints for tangible gestures in augmented reality,” in *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2007, pp. 1–4.

[3] D. Freeman, H. Benko, M. R. Morris, and D. Wigdor, “ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures,” in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, 2009, pp. 165–172.

[4] R. Sodhi, H. Benko, and A. Wilson, “LightGuide: projected visualizations for hand movement guidance,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 179–188.

[5] O. Bau and W. E. Mackay, “OctoPocus: a dynamic guide for learning gesture-based command sets,” in *Proceedings of the 21st annual ACM symposium on User interface software and technology*, 2008, pp. 37–46.

[6] J. Cassell, T. Bickmore, H. Vilhjálmsson, and H. Yan, “More than just a pretty face: affordances of embodiment,” in *Proceedings of the 5th international conference on Intelligent user interfaces*, 2000, pp. 52–59.

[7] P. Maes, T. Darrell, B. Blumberg, and A. Pentland, “The ALIVE system: Wireless, full-body interaction with autonomous agents,” *Multimed. Syst.*, vol. 5, no. 2, pp. 105–112, 1997.

[8] R. T. Azuma and others, “A survey of augmented reality,” *Presence*, vol. 6, no. 4, pp. 355–385, 1997.

[9] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, “Augmented reality technologies, systems and applications,” *Multimed. Tools Appl.*, vol. 51, no. 1, pp. 341–377, Jan. 2011.

[10] H. M. L, *Sensorama simulator*. Google Patents, 1962.

[11] I. E. Sutherland, “A head-mounted three dimensional display,” in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968, pp. 757–764.

[12] T. P. Caudell and D. W. Mizell, “Augmented reality: An application of heads-up display technology to manual manufacturing processes,” in *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*, 1992, vol. 2, pp. 659–669.

[13] L. B. Rosenberg, “‘Virtual Fixtures’: Perceptual Overlays Enhance Operator Performance in Telepresence Tasks,” Stanford University, Stanford, CA, USA, 1994.

[14] S. Feiner, B. Macintyre, and D. Seligmann, “Knowledge-based Augmented Reality.” 1993.

[15] H. KATO, “Artoolkit,” *http://www.hitl.washington.edu/artoolkit/*.

[16] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, and W. Piekarski, “ARQuake: An outdoor/indoor augmented reality first person application,” in *Wearable computers, the fourth international symposium on*, 2000, pp. 139–146.

[17] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari, “Multimodal human discourse: gesture and speech,” *ACM Trans. Comput.-Hum. Interact. TOCHI*, vol. 9, no. 3, pp. 171–193, 2002.

[18] T. Blum, V. Kleeberger, C. Bichlmeier, and N. Navab, “mirracle: An augmented reality magic mirror system for anatomy education,” in *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*, 2012, pp. 115–116.

[19] L. Fretwell, “StyleMeEngagementOverview_120611FINAL.pdf.” 2011.

[20] D. Norman, “DesignOfEverydayThings_text.pdf.” 1988.

[21] Victor Captelinin, “Affordance. In the Encyclopedia of Human Computer Interaction.” .

[22] R. Hartson, “Cognitive, physical, sensory, and functional affordances in interaction design,” *Behav. Inf. Technol.*, vol. 22, no. 5, pp. 315–338, Sep. 2003.

[23] A. Bragdon, R. Zeleznik, B. Williamson, T. Miller, and J. J. LaViola Jr, “GestureBar: improving the approachability of gesture-based interfaces,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 2269–2278.

[24] W. Delamare, C. Coutrix, and L. Nigay, “A tool for optimizing the use of a large design space for gesture guiding systems,” 2015, pp. 238–241.

[25] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice, “YouMove: enhancing movement training with an augmented reality mirror,” in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, 2013, pp. 311–320.

[26] B. Laurel, “Human Values and the Design of Computer Technology,” *Batya Friedman*, 1990. .

[27] H. Lieberman and T. Selker, *Agents for the user interface*. MIT Press, 2003.

[28] B. Shneiderman and P. Maes, “Direct manipulation vs. interface agents,” *interactions*, vol. 4, no. 6, pp. 42–61, 1997.

[29] P. Maes and others, “Agents that reduce work and information overload,” *Commun. ACM*, vol. 37, no. 7, pp. 30–40, 1994.

[30] H. Lieberman, “Autonomous interface agents,” in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 1997, pp. 67–74.

[31] P. Maes, “Artificial life meets entertainment: lifelike autonomous agents,” *Commun. ACM*, vol. 38, no. 11, pp. 108–114, 1995.

[32] H. Lieberman and others, “Letizia: An agent that assists web browsing,” *IJCAI 1*, vol. 1995, pp. 924–929, 1995.

[33] T. Rist, E. André, and J. Müller, “Adding animated presentation agents to the interface,” in *Proceedings of the 2nd international conference on Intelligent user interfaces*, 1997, pp. 79–86.

[34] J. C. Lester, S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. S. Bhogal, “The persona effect: affective impact of animated pedagogical agents,” in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 1997, pp. 359–366.

[35] B. Hartmann, M. Mancini, and C. Pelachaud, “Implementing expressive gesture synthesis for embodied conversational agents,” in *gesture in human-Computer Interaction and Simulation*, Springer, 2006, pp. 188–199.

[36] B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud, “Design and evaluation of expressive gesture synthesis for embodied conversational agents,” in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, 2005, pp. 1095–1096.

[37] S. C. Tan and A. Nareyek, “Integrating facial, gesture, and posture emotion expression for a 3D virtual agent,” in *Proceedings of the 14th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games (CGames 2009 USA)*, 2009, pp. 23–31.

[38] J. C. Lester, J. L. Voerman, S. G. Towns, and C. B. Callaway, “Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents,” *Appl. Artif. Intell.*, vol. 13, no. 4–5, pp. 383–414, May 1999.

[39] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan, “More than just a pretty face: conversational protocols and the affordances of embodiment,” *Knowl.-Based Syst.*, vol. 14, no. 1–2, pp. 55–64, Mar. 2001.

[40] C. Morris, M. E. Eiiwood, S. D. Mauk, and D. L. Hanley Jr, “Usability Study Analysis Report for,” pp. 189–194, 2001.

[41] “Measuring Usability with the System Usability Scale (SUS): MeasuringU.” [Online]. Available: http://www.measuringu.com/sus.php. [Accessed: 12-Jul-2016].

[42] H. L. O’Brien and E. G. Toms, *The development and evaluation of a survey to measure user engagement in e-commerce environments*. Wiley, 2010.

Appendix A. Focus Group User Study Questionnaire

1. Age: _____

2. Gender: Female / Male

3. Do you use public information systems (e.g. library kiosks, museum installation, tour information guide)?

☐ Never

☐ Few times a year

☐ Few times a month

☐ Few times a week

☐ Everyday

4. If you had used before, what kind of public information displays did you use? (please give context such as when or where)

5. Please circle on a number based on how much you agree:
I tend to actively try using an interactive public information display if I find one.

0	1	2	3	4	5	6	7	8	9	10
Strongly Disagree					Neutral					Strongly Agree

6. Do you use motion gesture interfaces (e.g. Xbox Kinect games)?

☐ Never

☐ Few times a year

☐ Few times a month

☐ Few times a week

☐ Everyday

7. If you had used before, what kind of gesture interfaces did you use?

8. I am good at figuring out how to use gesture interfaces.

0	1	2	3	4	5	6	7	8	9	10
Strongly Disagree					Neutral					Strongly Agree

9. I love characters (in movies, games, animations or cartoons).

0	1	2	3	4	5	6	7	8	9	10
Strongly Disagree					Neutral					Strongly Agree

10. If you have one, your favorite character is? (Please state where is the character from. e.g. movie, game, etc.)

11. I get affected by products having my favorite character on.

0	1	2	3	4	5	6	7	8	9	10
Strongly Disagree					Neutral					Strongly Agree

Thank you very much!

Appendix B. Experimental Procedure Script

- Welcome. Thank you for participating.
- (Let the participant sit at the table.)

- The purpose of this study is to identify visual cues to help interacting with large screen displays with gestures.
- In the experiment, we will ask you to stand in front of the display and interact with it and give us feedback through questionnaire and interview.
- Please be minded that we will be also recording the experiment for further analysis.
- The results will be published in academic publications including a master's thesis, but all the results and collected information will be anonymised to ensure privacy.
- The overall procedure will take about 30 minutes, and you can stop if you do not feel comfortable.
- Please read the information sheet and the consent form about the detail and if you agree to participate please sign the consent form. You can take the information sheet after the experiment if you want.
  Please let me know if you have any questions.
  (Leave the participant alone until done.)
- To begin the experiment, please fill in this questionnaire asking about your background.
- Please let me know if you have any questions.
  (Leave the participant alone until done. – prepare for trial 0)

  - The purpose of this study is to identify visual cues to help interacting with large screen displays with gestures.
  - There is nothing right or wrong with what you do.
  - We want to investigate how to improve the system so people can use the system in a more easy and intuitive way.

- (Start recording)

  - For the first trial, please stand in front of the screen and try to interact with it based on what is shown on the screen. (No hints given as this is the purpose).
- (Observe the following behaviours)
  How long did s/he took for figuring out what to do?
  How many errors they made for selecting the start button?

  - Please come and sit at the table and answer to the questionnaire.
  - Let me know if you are uncertain about answering the questions.

(Leave the participant alone until done)

- I’d like to ask some questions.
  (Ask the first interview questions)

- Now we are going to let you use the system for answering quiz on “robots and birds”.
  This is not about you neither answering the correct answer for the quiz nor getting the highest score. We don’t measure your score.
  The focus is on if the system is well designed to be used by novice users.
  Especially on how the visual cues shown on the screen help people interacting with it.

  You will be shown 3 different visual designs of the system.
  You will be asked to try them each, and answer a questionnaire after trying each of them.

  (Make sure to change the display condition in advance)
  (Go through the three conditions, trying out then answering per-task questionnaire,)
  (Check questionnaire if all questions are answered)

- Did you notice the difference between the three conditions?
  (If they are unsure, explain. PLAIN, TIMER, CHARACTER)
- Now based on all three conditions you tried. Please answer this questionnaire.
  (Leave the participant alone until done)

- I’d like to ask few more questions.
  (Ask the second interview questions).
  (Ask if the participant wants to review the interview scripts.)

- Thank you very much for your participation.
- Here is a small gift for participating.

Appendix C. Participant Information Sheet

PARTICIPANT INFORMATION SHEET

RESEARCH STUDY: **Study on visual affordance for natural gesture interaction with large screen displays**

RESEARCHERS: **Omprakash Rudhru, Dr. Gun Lee, Prof. Mark Billinghurst**

INTRODUCTION

You are invited to take part in a user interface design research study. Before you decide to be part of this study, you need to understand the risks and benefits. This information sheet provides information about the research study. The researcher will be available to answer your questions and provide further explanations. If you agree to take part in the research study, you will be asked to sign the consent form.

PURPOSE

The purpose of this study is to identify visual cues to help interacting with large screen displays with gestures.

PROCEDURE

The study will follow the procedure outlined as below:

1. The participant reads information sheet and signs the consent form.

2. The participant answers to a questionnaire on demographic information and his/her previous experience with using computer interfaces.

3. The researcher explains the study setup and experimental tasks for the participant to perform during the study.

4. The participant performs the experimental tasks including:

- Stand in front of a large screen display and watch the information on the screen.
- Perform the task following the information provided on the screen.
- Answer to a questionnaire asking for feedback on the given user interface.
*** The participant repeats the tasks above for the provided set of interfaces.**

5. The participant answers to a questionnaire asking for feedback on the overall study.

6. The participant will be asked a few questions for a debriefing interview.

The whole procedure will take approximately 30 minutes.

RISKS/DISCOMFORTS

Risks are minimal in this study. As you will be asked to act out gestures standing in front of a large display screen, it is expected that the experiment will involve physical movement of your body which could cause you feel tired or uncomfortable. However, as the level of physical activity will be within the range of everyday life activities, we do not expect any injury to come upon any of the participants.

CONFIDENTIALITY

All data obtained from participants will be kept confidential. In publications (e.g. Thesis, a public document which will be available through the UC Library), we will mainly report the results in an aggregate format: reporting only combined results and never reporting individual ones. In case of reporting quotes of the participants from the questionnaires, we will keep the source anonymous. Video of the experiment will be recorded for analysis purposes. All recordings will be concealed, and none other than the researchers will have access to them. If used for publication the certain parts of the body (e.g. face) will be hidden to keep anonymity. The data will be kept securely for up to 5 years and will be destroyed after completion of the research project.

PARTICIPATION

Participation in this research study is completely voluntary. You have the right to withdraw at any time or refuse to participate entirely.

COMPENSATION

Upon completion of participation in the study, the participant will receive a gift voucher.

APPROVAL OF THIS STUDY

This study has been reviewed and approved by the Human Interface Technology (HIT Lab NZ) and the University of Canterbury Human Ethics Committee Low Risk Approval process.

QUESTIONS

If you have questions or complaints regarding this study, please contact the researchers below or the University of Canterbury Human Ethics Committee.

Omprakash Rudhru (omprakash.rudhru@pg.canterbury.ac.nz)
Dr. Gun Lee (gun.lee@canterbury.ac.nz)
Prof. Mark Billinghurst (mark.billinghurst@canterbury.ac.nz)

University of Canterbury Human Ethics Committee
e-mail: human-ethics@canterbury.ac.nz

Please take this information sheet with you when you leave.

Appendix D. Participant Consent Form

PARTICIPANT CONSENT FORM

RESEARCH STUDY: **Study on visual affordance for natural gesture interaction with large screen displays**

RESEARCHER: **Omprakash Rudhru (omprakash.rudhru@pg.canterbury.ac.nz)**

SUPERVISORS: **Dr. Gun Lee (gun.lee@canterbury.ac.nz) , Prof. Mark Billinghurst (mark.billinghurst@canterbury.ac.nz)**

I have been given a full explanation of this project and have had the opportunity to ask questions. I understand what is required of me if I agree to take part in the research.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that any information or opinions I provide will be kept confidential to the researcher and the administrators of the research project and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

I understand that all data collected and recorded for the study will be kept in locked and secure facilities and/or in password protected electronic form up to five years and will be destroyed after completion of the research project.

I understand the risks associated with taking part and how they will be managed.

I understand that I am able to receive a report on the findings of the study by contacting the researcher at the conclusion of the project.

I understand that I can contact the researchers or supervisors listed above. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (e-mail: human-ethics@canterbury.ac.nz).

By signing below, I agree to participate in this research project, and I authorize recordings or other materials taken from this study used for scientific purposes, and I consent to publication of the results of the study.

Participant (Print name)	Signature	Date

Appendix E. Pre-Experiment Questionnaire

Age: _____________ ☐ Male / ☐ Female

Please check on ONE answer, unless it is described otherwise.

Do you use public information systems (e.g. library kiosks, museum installation, tour information guide)?

☐ Not at all

☐ A few times a year

☐ A few times a month

☐ A few times a week

☐ Every day

Have you used free hand gesture based interface before?

☐ Not at all

☐ A few times a year

☐ A few times a month

☐ A few times a week

☐ Every day

Have you played XBOX Kinect motion games before?

- ☐ Not at all
- ☐ A few times a year
- ☐ A few times a month
- ☐ A few times a week
- ☐ Every day

Have you played Nintendo Wii or Sony MOVE motion games before?

- ☐ Not at all
- ☐ A few times a year
- ☐ A few times a month
- ☐ A few times a week
- ☐ Every day

Please list all the motion game interfaces (e.g. Xbox Kinect) you have tried before.

Have you used an Augmented Reality (AR) app/interface before?

- ☐ I am not aware of what AR is.
- ☐ Not at all
- ☐ A few times a year
- ☐ A few times a month
- ☐ A few times a week
- ☐ Every day

Which hand do you usually use for pointing or making gestures?

- ☐ Left
- ☐ Right
- ☐ Both

Please check on the box based on how much you agree with each statement.

I consider myself using gestures a lot in everyday life.

Strongly disagree 1	2	3	Neutral 4	5	6	Strongly agree 7
☐	☐	☐	☐	☐	☐	☐

I love characters (in movies, games, animations or cartoons).

Strongly disagree 1	2	3	Neutral 4	5	6	Strongly agree 7
☐	☐	☐	☐	☐	☐	☐

I get affected by products having my favorite character on.

Strongly disagree 1	2	3	Neutral 4	5	6	Strongly agree 7
☐	☐	☐	☐	☐	☐	☐

Thank you! Please wait for further instruction.

Appendix F. Per-Trial Questionnaire

Please check on a box based on how much you agree with each statement.

Statement	Strongly disagree 1	2	Neutral 3	4	Strongly Agree 5
I think that I would like to use this system frequently.	☐	☐	☐	☐	☐
I found the system unnecessarily complex.	☐	☐	☐	☐	☐
I thought the system was easy to use.	☐	☐	☐	☐	☐
I think that I would need the support of a technical person to be able to use this system.	☐	☐	☐	☐	☐
I found the various functions in this system were well integrated.	☐	☐	☐	☐	☐
I thought there was too much inconsistency in this system.	☐	☐	☐	☐	☐
I would imagine that most people would learn to use this system very quickly.	☐	☐	☐	☐	☐
I found the system very cumbersome to use.	☐	☐	☐	☐	☐
I felt very confident using the system.	☐	☐	☐	☐	☐
I needed to learn a lot of things before I could get going with this system.	☐	☐	☐	☐	☐

Statement	Strongly disagree 1	2	Neutral 3	4	Strongly Agree 5
I would recommend this to my friends and family.	☐	☐	☐	☐	☐
I felt in control of my experience.	☐	☐	☐	☐	☐
I felt discouraged while using the interface.	☐	☐	☐	☐	☐
I could not do some of the things I needed to do on this interface.	☐	☐	☐	☐	☐
When I was doing the task, I lost track of the world around me.	☐	☐	☐	☐	☐
I lost myself in this experience.	☐	☐	☐	☐	☐
I was really drawn into my task.	☐	☐	☐	☐	☐
This experience was demanding.	☐	☐	☐	☐	☐
I blocked out things around me when I was doing the task.	☐	☐	☐	☐	☐
I felt interested in my task.	☐	☐	☐	☐	☐
I consider my experience a success.	☐	☐	☐	☐	☐
I felt involved in this task.	☐	☐	☐	☐	☐
This interface is attractive.	☐	☐	☐	☐	☐
My experience was rewarding.	☐	☐	☐	☐	☐
This experience did not work out the way I had planned.	☐	☐	☐	☐	☐
The experience was worthwhile.	☐	☐	☐	☐	☐

Statement	Strongly disagree 1	2	Neutral 3	4	Strongly Agree 5
Using this interface was mentally taxing.	☐	☐	☐	☐	☐
This interface was aesthetically appealing.	☐	☐	☐	☐	☐
The screen layout was visually pleasing.	☐	☐	☐	☐	☐
The time I spent just slipped away.	☐	☐	☐	☐	☐
I liked the graphics and images used.	☐	☐	☐	☐	☐
The content incited my curiosity.	☐	☐	☐	☐	☐
This interface appealed to my visual senses.	☐	☐	☐	☐	☐
I was so involved in my task that I lost track of time.	☐	☐	☐	☐	☐
I found this interface confusing to use.	☐	☐	☐	☐	☐
I felt annoyed while using the interface.	☐	☐	☐	☐	☐
This experience was fun.	☐	☐	☐	☐	☐
I was absorbed in my task.	☐	☐	☐	☐	☐
I continued to do the task out of curiosity.	☐	☐	☐	☐	☐
During this experience I let myself go.	☐	☐	☐	☐	☐
I felt frustrated while using the interface.	☐	☐	☐	☐	☐

Appendix G. Post-Experiment Questionnaire

Please check on the box based on how much you agree with each statement.

1. Please rank the condition based on your preference
Conditions: plain, with timer, with character

Best	1st place	
	2nd place	
Worst	3rd place	

What did you like of the condition you ranked in the 1st place?

What did you dislike of the condition you ranked in the 3rd place?

What could be improved in each condition?

Plain:

With timer:

With character:

Any comments on the overall user study?

Thank you for your participation!

Appendix H. Per-Task Questionnaire

Please answer in a scale of 0 to 10, 0: totally disagree, 10: totally agree, 5 is neutral.

2. Rate on a scale of 0 to 10 if it was easy to guess what to do at the first place, when you stood in front of the display.

3. Was it easy to know where to stand? (Rate in 0 ~10, then explain)

4. Any suggestions?

Appendix I. Post-experimental Questionnaire

Please check on the box based on how much you agree with each statement.

2. Please rank the condition based on your preference

Conditions: with timer, with character

Best	1st place	
	2nd place	

3. What did you like of the condition you ranked in the 1st place?

4. What did you dislike of the condition you ranked in the 2nd place?

5. What could be improved in each condition?

With timer:

With character:

6. Any comments on the overall user study?

Thank you for your participation!

Appendix J. Debriefing Interview

0. (Skim through the post experiment questionnaire answers and ask if hard to read.)

1. Did you felt if one of the conditions needed to wait longer to get the button selected (triggered)?

1-1 If yes, which one was the slowest/fastest?

2. Was the character attractive? Please rate with 0: Fully disagree ~ 10: Fully agree

3. If the character were more or less attractive, would it have affected your preference ranking?
(E.g. Yes/No/not sure/depends.)

3-1. (optional) if yes, how would you rank if the character was more or less attractive?
If more:
If less:

4. Which interface (Circular buttons / Bird) was the easiest to understand

5. Which interface (Circular buttons/ Bird) was more fun?

5. Any other comments?

6. Would you like to review the interview script?
[Observation Notes]

Appendix K. Observations

[Session 0]

- Smile on face?

□Yes □No □Not sure

- Trouble to find where to stand?

□Yes □No □Not sure

If yes, any details? (E.g. stood waiting for a while, confused with depth cue or which direction to move)

- Other observations:

[Session 1]

- How many and what kind of wrong gestures they made for selecting the circular button?
(E.g. tried pressing 2~3 times, grabbing fist, etc.)

- How many and what kind of wrong gestures they made for selecting the Bird button?
(E.g. tried pressing 2~3 times, grabbing fist, etc.)

- Which hand used for selecting the buttons?

□Left only □Right only □Both

- Any actions intentionally done (like answering wrong or moving the cursor etc.)

- Other observations:

Appendix L. Experimental Procedure Script

- Welcome. Thank you for participating.
- (Let the participant sit at the table.)

- The purpose of this study is to identify visual cues to help interacting with large screen displays with gestures.
- In the experiment, we will ask you to stand in front of the display and interact with it and give us feedback through questionnaire and interview.
  Please be minded that we will be also recording the experiment for further analysis.
  The results will be published in academic publications including a master's thesis, but all the results and collected information will be anonymised to ensure privacy.
  The overall procedure will take about 30 minutes, and you can stop if you do not feel comfortable.
- Please read the information sheet and the consent form about the detail and if you agree to participate please sign the consent form. You can take the information sheet after the experiment if you want.
  Please let me know if you have any questions.
  (Leave the participant alone until done.)
- To begin the experiment, please fill in this questionnaire asking about your background.
  Please let me know if you have any questions.
  (Leave the participant alone until done. – prepare for trial 0)

  - The purpose of this study is to identify visual cues to help interacting with large screen displays with gestures.
  - There is nothing right or wrong with what you do.
  - We want to investigate how to improve the system so people can use the system in a more easy and intuitive way.

- (Start recording)

  - For the first trial, please stand in front of the screen and try to interact with it based on what is shown on the screen. (No hints given as this is the purpose).
- (Observe the following behaviours)
  How long did s/he took for figuring out what to do?
  How many errors they made for selecting the start button?

  - Please come and sit at the table and answer to the questionnaire.
  - Let me know if you are uncertain about answering the questions.

  (Leave the participant alone until done)

- Now we are going to let you use the system for answering quiz on “Fruits and vegetables”.
  This is not about you. It is neither answering right nor getting the highest score. We don’t measure your score.
  The focus is on if the system is well designed to be used by novice users.
  Especially on how the visual cues shown on the screen help people interacting with it.

  You will be shown 2 different visual designs of the system.
  You will be asked to try them each, and answer a questionnaire after trying each of them.

  (Make sure to change the display condition in advance)
  (Go through the three conditions, trying out then answering per-task questionnaire,)
  (Check questionnaire if all questions are answered)

- Did you notice the difference between the three conditions?
  (If they are unsure, explain. TIMER, CHARACTER)
- Now based on all three conditions you tried. Please answer this questionnaire.
  (Leave the participant alone until done)

- I’d like to ask few more questions.
  (Ask the second interview questions).
  (Ask if the participant wants to review the interview scripts.)

- Thank you very much for your participation.
- Here is a small gift for participating.